Multivariate extreme value statistics for risk assessment

He, Yi

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Multivariate Extreme Value Statistics for Risk Assessment

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op dinsdag 6 december 2016 om 16.00 uur door

Yi He

geboren op 19 maart 1989 te Guangdong, China.

Promotores:	Prof. dr. John H.J. Einmahl
	Prof. dr. Bas J.M. Werker
Overige Commissie Leden:	dr. Pavel Čížek
	Prof. dr. Laurens de Haan
	Prof. dr. Davy Paindaveine
	dr. Chen Zhou

Acknowledgments

Three years ago I found my path back to Tilburg from Cambridge and started to work with Prof. John H.J. Einmahl in extreme value statistics. Upon finishing this dissertation, I have just realized all over again how fast time flies. Without the support from many people, especially John, it would not have been possible for me to complete these works.

I owe John a debt of endless gratitude for his kindness, generosity, and solicitude. John is a responsible supervisor, always open for advice, and an excellent teacher who can explain profound theories in simple language. He has also been a kind friend and an inspiring mentor, who keeps encouraging me to follow my academic passions and reminding me of a rigorous attitude toward science. There were countless times that I directly rushed into his office for all kinds of questions, and we sat down for hours of discussions. I have come up with many research ideas during the last three years; most of them were not that successful as those in this dissertation. John has always been listening patiently to each of them and has come back with detailed feedback and comments.

I also would like to express my gratitude to my second advisor, Prof. Bas Werker, who can always bring new thinking and give honest advice at the right time. I sincerely appreciate many of his comments from a practical perspective.

Many thanks to Prof. Liang Peng for his hospitality during my visit at Georgia State University and his support for my job search. It was a pleasure to work with him.

It is a great honor for me to have Prof. Laurens de Haan, Prof. Davy Paindaveine, Dr. Chen Zhou, and Dr. Pavel Cizek as my Ph.D. committee. Their comments and suggestions during my pre-defence were very helpful and led to this improved final version of my dissertation.

The Department of EOR and the Department of Finance in Tilburg University have offered me a great work environment and maximal academic freedom. I would like to thank Prof. Bertrand Melenberg and Otilia Boldea for their continuous help and support on my job search. I am grateful to Heidi, Lenie, Korine, and Anja for their excellent services and to Prof. Luc Renneboog for his advice at the very beginning of my Ph.D. study. I would like to thank Yanxi Hou, Andrea Krajina, and Juan-Juan Cai for their input and help on this dissertation.

I am thankful to many colleagues and friends who have made my life in the Netherlands full of fun and inspiration: Hailong Bao, Erwin Charlier, Feico Drost, Sebastian Ebert, Elsen Ho, Jan Kabatek, Xu Lang, Hong Li, Zhengyu Li, Alvin Lv, Gert Nieuwenhuis, Geng Niu, Henk Norde, Hans Schumacher, Johan Segers, Lei Shu, Oliver Spalt, Mitja Stadje, Yifan Yu, Nan Zhao, Bo Zhou, and many others.

Finally, my deepest thanks and love to my wife Xiaoyu and my parents for their understanding, patience, and caring. You are always by my side and supportive of my decision to pursue an academic career. It is because of you that I find the goodness beyond logic and the eternity behind all the randomness.

Tilburg, The Netherlands
September, 2016

Contents

Chapter 1

Introduction

Extremal events occur with small probability but often lead to catastrophic consequences. In economics, these events often involve large movements of asset prices, substantial losses or other tail behaviors of economic variables. The adverse effect associated with these so-called heavy-tail phenomena is often referred as the tail risk in financial econometrics.

This dissertation consists of three essays about statistical estimation and inference methods concerning extremal events and tail risks. Statistics of extremes is challenging because the tail behavior of economic variables is often governed by a very different law than that of its mean or median. While parametric methods can easily suffer from misspecification problems, fully non-parametric approaches often perform poorly due to the scarcity of extremal observations. Extreme value statistics adopt natural semi-parametric estimators from a coherent probabilistic theory of the sample maxima which is comparable to the theory of sums of random variables. Specifically, suppose we have a random sample $X_1, \ldots, X_n$ of some univariate positive-valued random variable X and for some sequence (a_n, b_n), we have

$$\frac{\max_{1 \leq i \leq n} X_i - b_n}{a_n} \xrightarrow{d} Y, \tag{1.0.1}$$

where Y is non-degenerate. It turns out the distribution of Y, subject to an appropriate affine transformation, is determined by a single parameter γ

called the *extreme value index*. Precisely, the distribution of Y is $G_\gamma(a\cdot+b)$ with $a>0$, b real and

$$G_\gamma(x)=\exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad 1+\gamma x>0.$$

with γ real and where for $\gamma=0$ the right-hand side is interpreted as $\exp(-e^{-x})$. When $\gamma<0$ there is a finite right endpoint of the support of X. When $\gamma=0$, all the moments of X exist and in this case we say the tail of X is light. When $\gamma>0$, we say X has a heavy tail and condition (1.0.1) is equivalent to the regular variation of the underlying distribution, that is,

$$\lim_{t\to\infty}\frac{\mathbb{P}(X>tx)}{\mathbb{P}(X>t)}=x^{-1/\gamma}, \quad x>0. \tag{1.0.2}$$

In this case, we can show that $E(X^r)=\infty$ if $r>1/\gamma$. The extreme value index of daily returns/losses of stocks, market indices, and exchange rates are often found to be between 0.2 and 0.4 in the finance literature; see, e.g., the survey papers by Cont (2001) and Gabaix (2009).

One of the most successful applications of extreme value statistics in risk management is the estimation of the univariate *extreme quantile*

$$q_\alpha=\inf\{q:\mathbb{P}(X>q)\le\alpha\}$$

where α is a given, small number. A natural extreme-value estimator is firstly proposed by Weissman (1978) and its asymptotic theory is well developed in literature; see, e.g., Section 4.3 in de Haan and Ferreira (2006).

The contribution of the next two chapters is a *multivariate* generalization of both the estimation procedure and asymptotic theory for the extreme quantile in arbitrary dimensions. Chapters 2 and 3 share the same spirit in bridging the concepts of *data depth* and extreme value theory. Since there is no complete ordering in $\mathbb{R}^d$ with $d\ge 2$, the notions of a multivariate quantile are established via the so-called data depth functions. Denote the underlying random vector as $\mathbf{X}\in\mathbb{R}^d$ and its distribution as P. A *data depth* is a P-based function from $\mathbb{R}^d$ into $[0,\infty)$, denoted as $D(\cdot)=D(\cdot;P)$ that, ideally, is maximized at a relevant center (also called median) of the distribution and decreases along the ray to zero from that centre, and satisfies many desirable properties such as affine equivariance; see, e.g., Zuo and Serfling (2000a). The depth value measures the centrality of a data point: extremely low depth corresponds to a substantial outlyingness relative to the center of the distribution. Our probabilistic model is heavily based on a multivariate analogue of the regular variation condition (1.0.2) as follows: for some non-degenerate so-called exponent measure ν

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{X} \in tB)}{\mathbb{P}(\|\mathbf{X}\| > t)} = \nu(B) \tag{1.0.3}$$

for all Borel set B such that $\nu(\partial B) = 0$. The exponent measure ν fully characterizes the tail dependence structure and heaviness of the underlying distribution.

Chapter 2 starts from a particular depth example called the *half-space depth* (Tukey, 1975) given by

$$HD(\mathbf{x}) = \inf \{P(H) : \mathbf{x} \in H \in \mathcal{H}\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

where $\mathcal{H}$ is the collection of all closed half-spaces. The half-space depth is among the most popular choices in non-parametric studies since it naturally satisfies many appealing properties regardless the underlying distribution. We propose a natural, semi-parametric estimator of the extreme depth-based *quantile region* given by

$$\mathcal{Q} = \left\{\mathbf{x} \in \mathbb{R}^d : HD(\mathbf{x}) < \beta\right\}$$

such that $p = P\mathcal{Q}$ is a given, small probability. In the spirit of extreme-value statistics, a refined consistency result is provided in the asymptotic embedding that $p = p_n \to 0$ as $n \to \infty$. The good performance of our extreme estimator is clearly demonstrated in a simulation study. The extreme depth-based quantile region is a collection of the joint extremal events of multiple risks, which may relate to financial data corresponding to irregular market behaviors such as erroneous trades and financial crises. It is important for the risk manager to understand the diversifiability between multiple assets, and to evaluate the portfolio performance during the unlikely scenarios; see, e.g., McNeil and Smith (2012).

Chapter 3 extends this approach to various depth functions, and, furthermore, establishes an asymptotic approximation theory of what-we-called *(directed) logarithmic distance* between our estimated and true quantile region. Therefore, we can construct (conservative) *confidence sets* that asymptotically cover the quantile region $\mathcal{Q}$ or its complement (often called the *central region*), or both simultaneously, with (at least) a prespecified probability under weak regular variation conditions. For the half-space depth, it is clear that the multivariate asymptotic theory has a distinctive nature from the univariate one, in the sense that the *shape* estimation error of the quantile region plays a significant role in finite samples.

Chapter 4 develops a statistical inference theory of a recently proposed tail risk measure by using the jackknife re-sampling technique and the empirical likelihood method which avoid complicated estimation of the asymptotic limit. This tail risk measure, which will be called *relative risk* henceforth, is proposed in Agarwal et al. (2016) as follows: given a bivariate random variable (X, Y) representing losses on, e.g., individual and market portfolios respectively, the relative risk of X against Y at level $\alpha \in (0, 1)$ is given by

$$\rho_\alpha = \rho_\alpha(X, Y) = \mathbb{P}(F_1(X) > 1 - \alpha | F_2(Y) > 1 - \alpha) \frac{E(X | F_1(X) > 1 - \alpha)}{E(Y | F_2(Y) > 1 - \alpha)},$$

where F_1, F_2 are the marginal distribution functions of X and Y respectively. It encompasses two parts: while the first part can be viewed as a finite-level analogue of the tail dependence coefficients (Sibuya, 1959)

$$\lambda = \lim_{\alpha \downarrow 0} \mathbb{P}(F_1(X) > 1 - \alpha | F_2(Y) > 1 - \alpha)$$

capturing the tail co-movement between X and Y, the second part is the ratio of the *expected shortfalls* of X and Y at the same level α. Agarwal et al. (2016) finds that the relative risk affects the cross-sectional variation in hedge fund returns. Examining relative risk measures of individual US bank equity losses against the market loss on Standard and Poor 500 (S&P 500), we document some empirical evidences of that maintaining a market level, i.e. unit level, of bank-specific relative risks implies a minimal future national *financial instability* during the 2009-crisis period. For regulators who are interested in monitoring the relative risks of individual banks, we provide a jackknife empirical likelihood inference procedure based on the smoothed nonparametric estimation and a Wilks type of result. The good coverage probability of the resulting confidence intervals is clearly shown in a simulation study.

Chapter 2

Estimation of Extreme Depth-based Quantile Regions

[Based on joint work with John H.J. Einmahl *Estimation of Extreme Depth-based Quantile Regions*, Journal of the Royal Statistical Society, forthcoming.]

Abstract. Consider the extreme quantile region induced by the halfspace depth function *HD* of the form $\mathcal{Q} = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) \le \beta\}$, such that $P\mathcal{Q} = p$ for a given, very small $p > 0$. Since this involves extrapolation outside the data cloud, this region can hardly be estimated through a fully nonparametric procedure. Using Extreme Value Theory we construct a natural, semiparametric estimator of this quantile region and prove a refined consistency result. A simulation study clearly demonstrates the good performance of our estimator. We use the procedure for risk management by applying it to stock market returns.

Key words. Extreme value statistics, halfspace depth, multivariate quantile, outlier detection, rare event, tail dependence.

2.1 Introduction

The Depth-Outlyingness-Quantiles-Ranks paradigm of Serfling (2010) states that the concepts of depth and quantile are essentially equivalent under some regularity conditions for a $\mathbb{R}^d$-valued random vector, say $\boldsymbol{X}$. A statistical depth function (Definition 2.1 in Zuo and Serfling, 2000a) provides a probability based ordering from the center (the point with maximal depth value) outwards and therefore induces a multivariate quantile function and vice versa under suitable regularity conditions, see, e.g., Serfling (2006). Here we consider a seminal example introduced in Tukey (1975), called the halfspace depth $HD : \mathbb{R}^d \to [0, \infty)$ defined by

$$HD(\boldsymbol{x}, P) = \inf\{P(H) : \boldsymbol{x} \in H \in \mathcal{H}\}, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

where P is the probability measure of $\boldsymbol{X}$ and $\mathcal{H}$ is the class of closed halfspaces.

The depth function measures the outlyingness of points relative to the center from a global perspective. The *extreme* depth-based quantile region consists of the extremely outlying points, that is, it is of the form

$$\mathcal{Q} = \mathcal{Q}(\boldsymbol{X}, \beta) = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) \leq \beta\} \tag{2.1.1}$$

for a given, very small number $p = P\mathcal{Q} > 0$. (In the sequel, without confusion, we use the notations $\mathcal{Q}$, $\mathcal{Q}_{\boldsymbol{X}}$, $\mathcal{Q}(\boldsymbol{X}, \beta)$ and $\mathcal{Q}(\boldsymbol{X}; p)$ interchangeably.) It is the (closure of the) complement of the $(1 - p)$th central region which itself enjoys many desirable properties including convexity (if P has a continuous distribution function) and nestedness, see Zuo and Serfling (2000b). The extreme quantile contour is defined accordingly as $\mathcal{C} = \mathcal{C}_\beta = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) = \beta\}$. Remarkably, the quantile region can be also generated without any depth setting by the directional quantile, see Hallin et al. (2010) and Kong and Mizera (2012). However, these two approaches require an explicit value of β, which is unknown here in general. Extreme multivariate quantiles defined similarly but in terms of the probability density are studied in Cai et al. (2011), see Remark 2.2.5 below.

The extreme depth-based quantile has strong practical values, particularly in economics and finance studies. A direct application is to detect data outliers, which occur with extremely small probability, e.g. financial data corresponding to irregular market behavior such as erroneous trades and financial crises. A second application is to reveal the jointly extreme behavior of multivariate risks. This is important for the risk/portfolio manager to understand the diversifiability between multiple risks/assets. Last but not least, the extreme depth-based quantile can define the unlikely scenarios for stress testing (McNeil and Smith, 2012).

The purpose of this paper is to estimate the quantile region $\mathcal{Q}$ (or the quantile contour $\mathcal{C}$) from a random sample from P. A natural nonparametric estimator of $\mathcal{Q}$ can be obtained by simply exploiting the sample depth function. Here in the spirit of extreme value statistics, p is very small and typically of order $1/n$. This means that the number of data points that lie in $\mathcal{Q}$ is small and can even be zero, leaving little information for estimating it nonparametrically. Indeed, the estimator based directly on the sample depth will perform poorly, which is demonstrated clearly in our simulation study.

We consider multivariate regularly varying distributions since our interest is in extreme quantile regions that are far away from the distribution center and the origin; see, e.g., Section 5.4 in Resnick (2007).

Assumption 2.1.1. The random vector $\boldsymbol{X}$ is multivariate regularly varying, that is, there exists a measure ν (the exponent measure) such that, as $t \to \infty$,

$$\frac{\mathbb{P}(\boldsymbol{X} \in tB)}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \to \nu(B) < \infty \tag{2.1.2}$$

for every Borel set $B \subset \mathbb{R}^d$ that is bounded away from the origin and satisfies $\nu(\partial B) = 0$ and $tB = \{t\boldsymbol{x} : \boldsymbol{x} \in B\}$. In addition, let $\nu(B) > 0$ if $B \supset H$ for some $H \in \mathcal{H}$.

Here $\|\cdot\|$ can denote any norm on $\mathbb{R}^d$. For convenience, we take $\|\cdot\|$ as the L_2-norm throughout this paper. This limit relation is a multivariate analogue of the regular variation condition in univariate extreme value theory, when the probability distribution is in the max domain of attraction of a Fréchet distribution. It is satisfied by many multivariate distributions with a heavier tail. Examples include those in the sum-domain of attraction of α-stable distributions and elliptical distributions with heavy tails such as multivariate t-distributions. When $d = 2$, it can also be tested formally using the procedure in Einmahl and Krajina (2016). It follows that ν is homogeneous, that is, there exists a $\gamma > 0$ such that for all $t > 0$

$$\nu(tB) = t^{-1/\gamma}\nu(B); \tag{2.1.3}$$

see, e.g., de Haan and Resnick (1979). The number γ is called the extreme value index. Clearly, ν defines a probability measure on the complement of the open unit ball in $\mathbb{R}^d$. Exploiting this assumption we will construct an estimator of $\mathcal{Q}$ based on the statistics of extremes methodology. We shall show that ν asymptotically determines the shape of extreme quantile region. We also assume that the measure ν is positive on halfspaces to prevent that the extreme quantile regions will be degenerate in some directions.

While there exist many different notions of data depth, the halfspace depth has many appealing intrinsic properties regardless the underlying distribution and broad applicability. It is therefore often preferred in nonparametric studies, see, for example, Donoho and Gasko (1992), Yeh and Singh (1997), Struyf and Rousseeuw (1999), and Liu et al. (1999). In the survey paper Zuo and Serfling (2000a) this is summarized by "it is found that the halfspace depth behaves very well overall in comparison with various competitors". Other depths, like the Mahalanobis (1936), spatial (Chaudhuri, 1996; Serfling, 2002) or projection-based depth (Zuo, 2003) are useful for many applications but are established mainly for their distributional characteristics in the central region but not in the tail. In contrast, as shown below, the halfspace depth conveys profound information about the probabilistic structure of the tail and provides a natural link to multivariate extreme value theory. More precisely we have the following: if $\tilde{\boldsymbol{X}}$ has probability measure $\widetilde{P}$ and P and $\widetilde{P}$ are identical outside some bounded subset of $\mathbb{R}^d$, then for the halfspace depth and very small p, $\mathcal{Q}(\boldsymbol{X}; p) = \mathcal{Q}(\tilde{\boldsymbol{X}}; p)$, whereas for one of the just mentioned depths (Mahalanobis, spatial, projection-based) we do not necessarily have $P(\mathcal{Q}(\boldsymbol{X}; p) \Delta \mathcal{Q}(\tilde{\boldsymbol{X}}; p))/p \to 0$ as $p \downarrow 0$ (where Δ denotes 'symmetric difference').

It is inconvenient that outside the convex hull of the data the sample halfspace depth is equal to 0. This could be circumvented by considering a smoothed version of the empirical distribution that is supported on the whole $\mathbb{R}^d$. Our proposed procedure can be seen as based on such a smoothed version of the empirical distribution in the tail, where the smoothing is done by using extreme value statistics. This has not only the advantages of smoothing the point masses and yielding positive values (which could be done in many ways), but, most importantly, it also yields a statistically much better estimator of the halfspace depth in the tail. Many other depths, e.g. the spatial depth, Mahalanobis depth and projection-based depth, do not suffer from the discreteness of the empirical distribution, but this in itself does not guarantee good statistical properties in the tail of their empirical versions. The estimation of their corresponding extreme quantile region remains an issue because of the unknown underlying depth value β, which is very difficult to approximate in the tail.

This paper is organized as follows. In Section 2 we construct our estimator and show some of its properties and we establish a refined consistency result. Section 3 demonstrates the excellent performance of our estimator in a simulation study while Section 4 presents a real-life financial application. The proofs are deferred to the end.

The data that are analysed in the paper and the programs that were used to analyse them may be obtained from http://wileyonlinelibrary.com/journal/rss-datasets

2.2 Main Results

Consider a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from P. Define the radii $R = \|\boldsymbol{X}\|$ and $R_i = \|\boldsymbol{X}_i\|$ for $i = 1, \ldots, n$. We order the R_i's as $R_{1:n} \leq \ldots \leq R_{n:n}$. Define $F_R(t) = \mathbb{P}(R \leq t)$ and $U(t) = F_R^{\leftarrow}\left(1 - \frac{1}{t}\right)$, where $F_R^{\leftarrow}$ is the left-continuous inverse of F_R. We require:

Assumption 2.2.1. $P(\mathcal{C}_\beta) = 0$ for all $\beta > 0$.

This is to ensure the existence of $\mathcal{Q}$ for all $p \in (0, 1)$.

Proposition 2.2.1. *Under Assumption 2.2.1, for any* $0 < p < 1$, *it holds that* $P(\mathcal{Q}(\boldsymbol{X}, \beta)) = p$, *where* $\beta = \sup\{\widetilde{\beta} : P(\mathcal{Q}(\boldsymbol{X}, \widetilde{\beta})) \leq p\}$.

It follows from above that the function $\mathbb{P}(R \geq t)$, $t > 0$, is regularly varying at infinity with exponent $-1/\gamma$. We further assume:

Assumption 2.2.2.

$$\lim_{t\to\infty} \frac{\mathbb{P}(R \geq t)}{t^{-1/\gamma}} = c \in (0, \infty).$$

This is weaker than the often used second-order condition with a negative second order parameter ρ, see Theorem 2.3.9 in de Haan and Ferreira (2006).

We parametrize the halfspace $H = H_{r,\boldsymbol{u}}$ by a pair of parameters $(r, \boldsymbol{u})$ with $r \in \mathbb{R}$ and $\boldsymbol{u} \in \Theta := \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| = 1\}$. Here $\boldsymbol{u}$ is its unit normal vector and r is the lower bound of the inner product between $\boldsymbol{u}$ and points in H. Precisely, we write

$$H_{r,\boldsymbol{u}} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{u}^T\boldsymbol{x} \geq r\}$$

and its collection $\mathcal{H} = \{H_{r,\boldsymbol{u}} : r \in \mathbb{R}, \boldsymbol{u} \in \Theta\}$. Then the halfspace depth function can be written in a simplified way as

$$HD(\boldsymbol{x}, P) = \inf_{\boldsymbol{u}\in\Theta} P(H_{\boldsymbol{u}^T\boldsymbol{x},\boldsymbol{u}}).$$

Therefore the extreme quantile region we wish to estimate can be rewritten as

$$\mathcal{Q} = \{\boldsymbol{x} \in \mathbb{R}^d : \inf_{\boldsymbol{u}\in\Theta} P(H_{\boldsymbol{u}^T\boldsymbol{x},\boldsymbol{u}}) \leq \beta\}$$

where $P\mathcal{Q} = p \in (0,1)$ with $p = p_n \to 0$ as $n \to \infty$. This means that both $\mathcal{Q}$ and β depend on n, that is $\mathcal{Q} = \mathcal{Q}_n$ and $\beta = \beta_n$.

Accordingly to Tukey's halfspace depth, define the *extreme* halfspace depth function by

$$HD(\boldsymbol{z},\nu) = \inf\{\nu(H) : \boldsymbol{z} \in H \in \mathcal{H}\} = \inf_{\boldsymbol{u}\in\Theta} \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}), \quad \boldsymbol{z} \neq (0,\ldots,0)^T.$$

Observe that $\nu(H_{r,\boldsymbol{u}}) = \infty$ for any halfspace $H_{r,\boldsymbol{u}}$ with $r \leq 0$.

There is a uniform limit relation, analogous to (2.1.2), between $HD(\cdot,P)$ and $HD(\cdot,\nu)$:

Proposition 2.2.2. *Under Assumptions 1 - 3, for any $\varepsilon > 0$,*

$$\lim_{t\to\infty} \sup_{\|\boldsymbol{z}\|\geq\varepsilon} \left| \frac{HD(t\boldsymbol{z},P)}{\mathbb{P}(R \geq t)} - HD(\boldsymbol{z},\nu) \right| = 0.$$

We derive our estimator by using this relation with $t = U(n/k)$, where $k = k_n \in \{1,\ldots,n\}$ is an intermediate sequence, that is,

Assumption 2.2.3. $k = k_n$ satisfies $k \to \infty$ and $k/n \to 0$, as $n \to \infty$.

The second part is needed to apply Proposition 2.2.2; the first part will ensure that the effective sample size tends to ∞. Now with Proposition 2.2.2 and the homogeneity property of ν, we can approximate $\mathcal{Q}$ with

$$\left\{U(n/k)\boldsymbol{x} \in \mathbb{R}^d : \frac{k}{n} HD(\boldsymbol{x},\nu) \leq \beta\right\} = U\left(\frac{n}{k}\right)\left(\frac{k}{n\beta}\right)^{\gamma} \{\boldsymbol{z} \in \mathbb{R}^d : HD(\boldsymbol{z},\nu) \leq 1\}. \tag{2.2.1}$$

Substituting the implicit β by its approximation $p/\nu(S)$, see Lemma 6 in the on-line supplementary material, yields that

$$\mathcal{Q} \approx U\left(\frac{n}{k}\right)\left(\frac{k\nu(S)}{np}\right)^{\gamma} S =: \widetilde{\mathcal{Q}}_n$$

where

$$S = \{\boldsymbol{z} \in \mathbb{R}^d : HD(\boldsymbol{z}, \nu) \leq 1\} = \{\boldsymbol{z} = r\boldsymbol{w} : r \geq (HD(\boldsymbol{w}, \nu))^\gamma, \boldsymbol{w} \in \Theta\}.$$

Hence we need estimators for $U(\frac{n}{k})$, γ, $\nu(S)$ and S. We start from $\widehat{U}\left(\frac{n}{k}\right) = R_{n-k:n}$, the $(k+1)$-st largest radius in the data. The extreme value index γ can be estimated using the univariate data of radii by various methods; see, e.g., Hill (1975), Smith (1987) and Dekkers et al. (1989). The typical convergence rate of the estimator $\widehat{\gamma} > 0$ is of order $k^{-1/2}$. For the rest, it is sufficient to provide an estimator of the measure ν, which determines both the set S and $\nu(S)$. A natural estimator of $\nu(B)$ on any Borel set B, which is away from the origin, is to use the sample version

$$\widehat{\nu}(B) = \frac{1}{k+1} \sum_{i=1}^{n} \mathbb{1}\left[\frac{\boldsymbol{X}_i}{R_{n-k:n}} \in B\right] = \frac{n}{k+1} P_n(R_{n-k:n} B),$$

where P_n is the empirical probability measure of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. However, to recover the homogeneity of ν in our estimation we adopt another estimator on halfspaces $H_{r,\boldsymbol{u}}$ given by $\widehat{\nu}^*(H_{r,\boldsymbol{u}}) = r_+^{-1/\widehat{\gamma}} \widehat{\nu}(H_{1,\boldsymbol{u}})$ with $r_+ = \max\{r, 0\}$. Then we define

$$\widehat{S} = \{\boldsymbol{z} = r\boldsymbol{w} : r \geq (HD(\boldsymbol{w}, \widehat{\nu}^*))^{\widehat{\gamma}}, \boldsymbol{w} \in \Theta\}.$$

Collecting all the estimators above we estimate $\mathcal{Q}_n$ by

$$\widehat{\mathcal{Q}}_n = \widehat{\mathcal{Q}}_n(\boldsymbol{X}; p) = \widehat{U}\left(\frac{n}{k}\right) \left(\frac{k\widehat{\nu}(\widehat{S})}{np}\right)^{\widehat{\gamma}} \widehat{S}$$

and $\mathcal{C}_n$ by

$$\widehat{\mathcal{C}}_n = \left\{ \widehat{U}\left(\frac{n}{k}\right) \left(\frac{k\widehat{\nu}(\widehat{S})}{np}\right)^{\widehat{\gamma}} (HD(\boldsymbol{w}, \widehat{\nu}^*))^{\widehat{\gamma}} \boldsymbol{w} : \boldsymbol{w} \in \Theta \right\}.$$

We present some properties of the estimated quantile region $\widehat{\mathcal{Q}}_n$.

Proposition 2.2.3. *Under Assumption 2.2.1, the estimated quantile regions have, almost surely, following properties:*

(a) The complement of $\widehat{\mathcal{Q}}_n$, denoted as $\widehat{\mathcal{Q}}_n^c$, is bounded and convex.

(b) Orthogonal and scale equivariance: for any orthogonal $d \times d$ matrix $\boldsymbol{R}$ and $c > 0$, provided the estimator $\widehat{\gamma}$ is (positive) scale invariant (e.g. Hill, 1975; Smith, 1987; Dekkers et al., 1989), it holds that

$$\widehat{\mathcal{Q}}_n(c\boldsymbol{R}\boldsymbol{X}; p) = c\boldsymbol{R}\widehat{\mathcal{Q}}_n(\boldsymbol{X}; p) := \{c\boldsymbol{R}\boldsymbol{x} : \boldsymbol{x} \in \widehat{\mathcal{Q}}_n(\boldsymbol{X}; p)\}.$$

(c) The $\widehat{\mathcal{Q}}_n$ are nested: for $p_1 < p_2$, $\widehat{\mathcal{Q}}_n(\boldsymbol{X}; p_1) \subset \widehat{\mathcal{Q}}_n(\boldsymbol{X}; p_2)$.

For similar results for quantile regions based on the true or sample half-space depth, see Donoho and Gasko (1992) and Zuo and Serfling (2000a, 2000b).

We now present our main result with "$\xrightarrow{\mathbb{P}}$" denoting convergence in probability.

Theorem 2.2.1. *Suppose Assumptions 1 - 4 hold and $\widehat{\gamma}$ is an estimator such that $\sqrt{k}(\widehat{\gamma} - \gamma) = O_{\mathbb{P}}(1)$. If, as $n \to \infty$, $(\log np)/\sqrt{k} \to 0$, then*

$$\sup_{\boldsymbol{x} \in \widehat{\mathcal{C}}_n} |\log HD(\boldsymbol{x}, P) - \log \beta| \xrightarrow{\mathbb{P}} 0 \quad \textit{and} \quad \frac{P(\widehat{\mathcal{Q}}_n \Delta \mathcal{Q})}{p} \xrightarrow{\mathbb{P}} 0.$$

Remark 2.2.1. The above approach treats p as explicitly given and solves the implicit β. We consider that it is also natural to, instead, have β explicitly given; see, e.g., Hallin et al. (2010) and Kong and Mizera (2012). In this case one step in the derivation of the estimator can be omitted: the replacement of β by its unknown asymptotic substitute $p/\nu(S)$ is not necessary now and hence the procedure becomes easier, see equation (2.2.1) and below. In particular we do not need to estimate $\nu(S)$. Precisely, the estimated region becomes

$$\widehat{\mathcal{Q}}_n^* = \widehat{U}\left(\frac{n}{k}\right)\left(\frac{k}{n\beta}\right)^{\widehat{\gamma}} \widehat{S}$$

and the modified quantile contour $\widehat{\mathcal{C}}_n^*$ can be defined analogously. Proposition 2.2.3 and Theorem 2.2.1 still hold with $\widehat{\mathcal{Q}}_n$ replaced by $\widehat{\mathcal{Q}}_n^*$ and $\widehat{\mathcal{C}}_n$ by $\widehat{\mathcal{C}}_n^*$.

Remark 2.2.2. When p is sufficiently small we can write $\partial\mathcal{Q} = \{\rho(\boldsymbol{w})\boldsymbol{w} : \boldsymbol{w} \in \Theta\}$ and $\widehat{\mathcal{C}}_n = \{\widehat{\rho}(\boldsymbol{w})\boldsymbol{w} : \boldsymbol{w} \in \Theta\}$ with (unique) positive radius functions ρ and $\widehat{\rho}$. Then using the intermediate results in the on-line supplementary material, it can be shown that

$$\sup_{\boldsymbol{w}\in\Theta}\left|\frac{\widehat{\rho}(\boldsymbol{w})}{\rho(\boldsymbol{w})} - 1\right| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{\lambda(\widehat{\mathcal{Q}}_n \Delta \mathcal{Q})}{\lambda(\mathcal{Q}_n^c)} \xrightarrow{\mathbb{P}} 0,$$

where λ denotes Lebesgue measure.

Remark 2.2.3. We can separate the choices of k for estimation of γ and the measure ν, respectively k_γ and k_ν, say. Then Theorem 1 requires that both k_γ and k_ν satisfy Assumption 2.2.3, $\sqrt{k_\gamma}(\widehat{\gamma}-\gamma) = O_{\mathbb{P}}(1)$, and $(\log \frac{k_\nu}{np})/\sqrt{k_\gamma} \to 0$.

The actual choice of k for a finite sample is a well-known issue. A heuristic guideline is to choose a k that gives almost the same estimates in its neighborhood. For example, here a two-step selection procedure may be adopted. Plot $\widehat{\gamma}$ against k, search for the first stable region in the graph and choose k_γ to be the midpoint of this region and find an estimate of γ. Then choose k_ν in a similar manner by plotting $\widehat{\nu}(\widehat{S})$ (using the just obtained $\widehat{\gamma}$) against k.

Remark 2.2.4. Note that $\widetilde{\mathcal{Q}}_n$ has the same shape as S, which does not depend on n. This means that the extreme quantile regions $\mathcal{Q} = \mathcal{Q}_n$ are approximately homothetic. Here the limiting shape, i.e. the shape of S, is fully characterized by the exponent measure ν. In general, the shape of the extreme quantile region is determined by the choice of the depth function but not necessarily by the *tail* of the distribution. For example, for the projection-based depth this shape is determined by the scale measure, which is usually taken to be the median absolute deviation (MAD) of the projection random variable, see Zuo (2003).

Remark 2.2.5. The aforementioned paper Cai et al. (2011) studies related extreme quantile regions defined in terms of the density instead of the depth. Therefore, in contrast to this paper, for the construction of those quantile regions obviously the existence of the density is needed and for deriving their asymptotic properties the stronger multivariate regular variation at the density level is required. Hence the present method has a broader applicability. Note that the density-based regions can be very different from the present ones, e.g., their corresponding central regions need not be convex. It depends on the type of application which features of the region are preferred.

Remark 2.2.6. In the recent paper Einmahl et al. (2015a) the sample halfspace depth has been refined to yield an estimator that performs well in both the central part of the distribution and the tail. The procedure and the goal of the present paper are substantially different from those of that paper. There the goal is to estimate $HD(\cdot, P)$ well on a very large region in $\mathbb{R}^d$ and to apply this refined estimator, whereas here we focus on a procedure that performs well in the tail and use it for estimating extreme quantile regions. More specifically, there the refinement of the estimator is done first at the univariate level for the projected data, whereas here directly a multivariate approach is used.

2.3 Simulation Study

In this section a simulation study is carried out to evaluate the finite-sample performance of our extreme quantile estimator. The extreme value index γ is estimated by the Hill (1975) estimator. Boxplots are presented based on 100 scenarios. We consider the following multivariate distributions.

- The bivariate Cauchy distribution ($\gamma = 1$) with density

$$f(x, y) = \frac{1}{2\pi(1 + x^2 + y^2)^{3/2}}, \quad (x, y) \in \mathbb{R}^2.$$

- The bivariate Student-t_3 distribution ($\gamma = 1/3$) with density

$$f(x, y) = \frac{1}{2\pi(1 + (x^2 + y^2)/3)^{5/2}}, \quad (x, y) \in \mathbb{R}^2. \tag{2.3.1}$$

- A bivariate elliptical distribution ($\gamma = 1/3$) with density

$$f(x,y) = \frac{3(x^2/4+y^2)^2}{4\pi\big(1+(x^2/4+y^2)^3\big)^{3/2}}, \quad (x,y) \in \mathbb{R}^2.$$

- An affine transformation of the bivariate Cauchy ($\gamma = 1$) random vector $\boldsymbol{Y}$:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{b}, \quad \boldsymbol{A} = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}. \tag{2.3.2}$$

- A bivariate "clover" distribution ($\gamma = 1/3$) with density

$$f(x,y) = \frac{3\Big(9(x^2+y^2)^2 - 32x^2y^2\Big)}{10\pi\big(1+(x^2+y^2)^3\big)^{3/2}}, \quad (x,y) \in \mathbb{R}^2.$$

  This is a distribution with clover-shaped (hence non-elliptical and non-convex) density contours; cf. Cai et al. (2011). Recall that, however, halfspace-depth based quantile contours are always convex.

- The trivariate Cauchy distribution ($\gamma = 1$) with density

$$f(x) = \frac{1}{\pi^2(1+x^2+y^2+z^2)^2}, \qquad (x,y,z) \in \mathbb{R}^3. \tag{2.3.3}$$

Figure 2.1 shows the true and estimated quantile regions of the bivariate distributions for $p = 1/2000$, $1/5000$, or $1/10000$ with sample size $n = 5000$ and $k = 400$. (For the bivariate clover distribution we can depict only approximate true quantile contours because of computational complexity.) The estimated regions are all close to the true ones. It is clear that our (estimated) extreme quantile regions belong to an 'almost empty' space, i.e., a space with few or even no observations.

Table 2.1 shows the median of the relative errors of our extreme (EVT) estimator for $p = 1/5000$ based on 100 samples of size $n = 5000$ or $n = 1000$. In the former case we consider three different choices of k: 200, 400 and 800. Our EVT estimator performs well for all these cases.

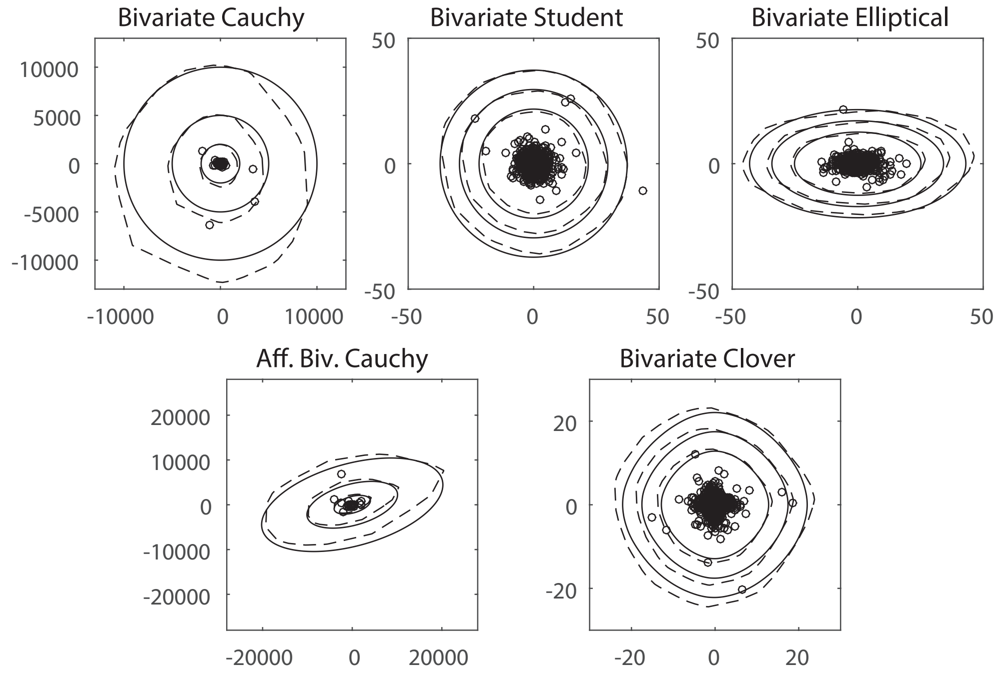


Figure 2.1: True (solid) and estimated (dashed) quantile regions for $p = 1/2000$, $1/5000$ or $1/10000$ based on one sample of size 5000 with choice of $k = 400$.

	$P(\widehat{\mathcal{Q}}_n \Delta \mathcal{Q})/p$				$\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} \lvert\log HD(\boldsymbol{x}, P) - \log\beta\rvert$			
Biv. Cauchy	0.35	0.21	0.22	0.43	0.49	0.34	0.30	0.62
Biv. Student t_3	0.42	0.29	0.33	0.42	0.70	0.52	0.55	0.84
Elliptical	0.37	0.26	0.20	0.64	0.77	0.53	0.39	1.06
Affine Cauchy	0.30	0.30	0.38	0.52	0.55	0.47	0.60	0.82
Triv. Cauchy	0.29	0.32	0.23	0.47	0.54	0.51	0.36	0.81

Table 2.1: Median of the relative errors of EVT estimates for $p = 1/5000$ based on 100 samples. In both panels, the first three columns in each panel are for $n = 5000$ with $k = 200, 400, 800$ and the last column is for $n = 1000$ with $k = 150$.

Next we compare our EVT estimator to the (fully) nonparametric estimators with $n = 5000$. Only the estimator of the depth, not the quantile,

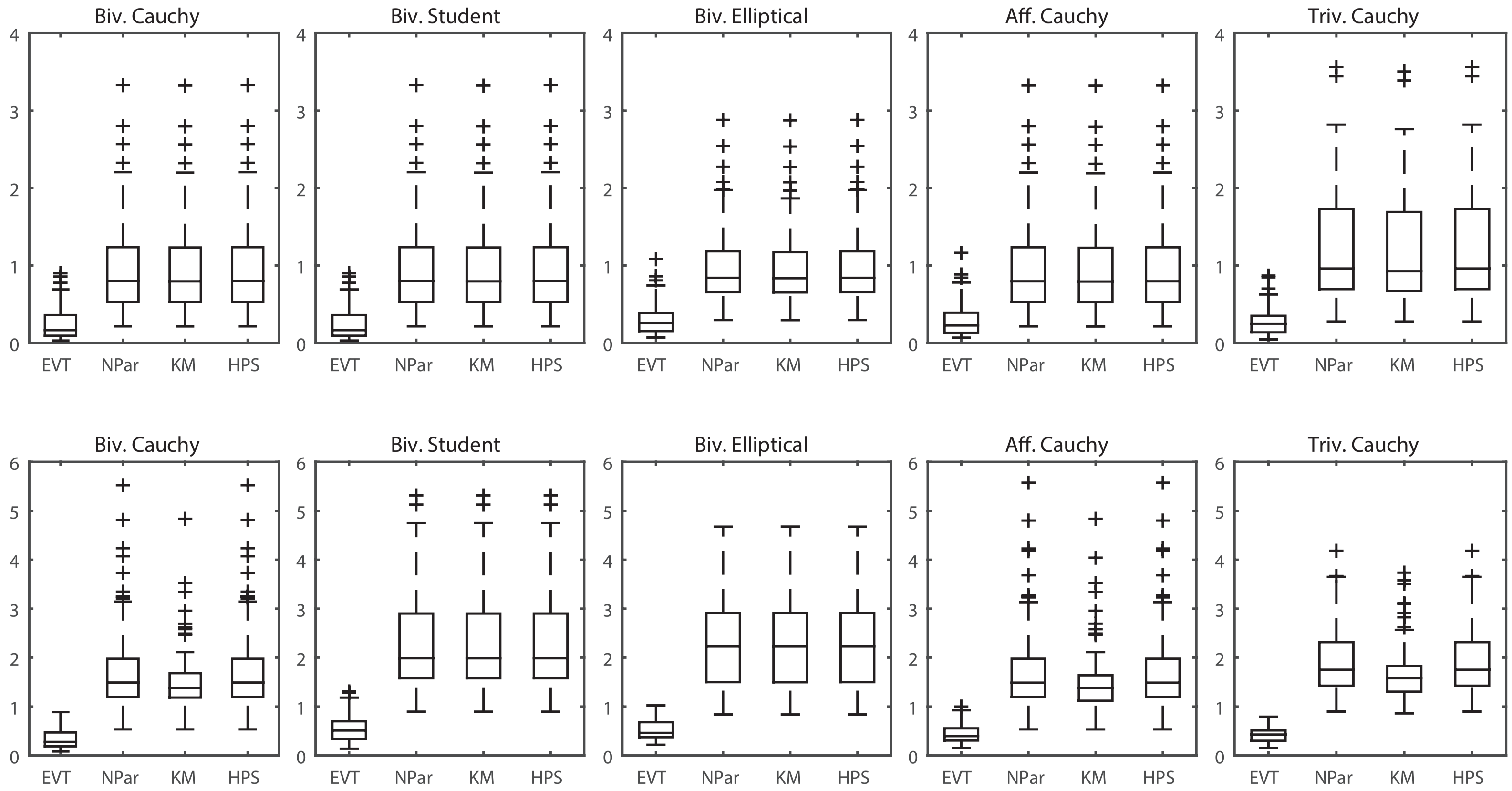


Figure 2.3: Boxplots of $P(\widehat{\mathcal{Q}}_n^* \Delta \mathcal{Q})/p$ (first row) and $\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n^*} |\log HD(\boldsymbol{x}, P) - \log\beta|$ (second row) for the extreme (EVT), nonparametric (NPar), KM (Kong and Mizera, 2012) and HPS (Hallin et al., 2010) estimates based on 100 samples of size 5000 with $\beta = 1/5000$. We choose $k = 400$ for the EVT estimates.

contours are established in the nonparametric literature. Therefore we consider the cases with $\beta = 1/n$ and use the modified estimator $\widehat{\mathcal{Q}}_n^*$ for the extreme quantile region, see Remark 1, to ensure these methods are comparable. A simple nonparametric estimator is the closure of the complement of the convex hull of the data, which is directly based on the sample depth function. Alternatively, the quantile regions can be estimated using the envelope of the sample directional quantile lines by Hallin et al. (2010) or Kong and Mizera (2012). Figure 2.2 shows an example. Clearly our EVT estimator completely outperforms the nonparametric ones.

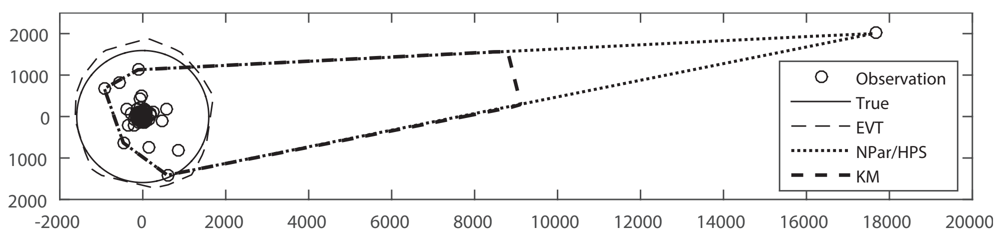


Figure 2.2: True and estimated quantile regions of bivariate Cauchy distribution for $\beta = 1/5000$ based on one sample of size 5000, for $k = 400$.

Figure 2.3 clearly demonstrates the good performance of the EVT estimator. It produces much smaller medians and ranges of relative errors at both the probability and depth level for all the distributions we consider compared to the fully nonparametric approaches.

2.4 Application

In this section we present a real-world finance application. The dataset, downloaded from Datastream, consists of the daily international market price indexes of Standard and Poors S&P 500 from the United States, the Financial Times Stock Exchange FTSE 100 from the United Kingdom and the Nikkei 225 index from Japan. The sample period is from July 2, 2001 to June 29,

2007. The daily market return is then computed as the logarithm of the ratio of current and one-period ago price, giving rise to 1564 observations for each country.

As usual the squared stock returns exhibit moderate autocorrelation and the Ljung-Box test rejects the serial independence for all these univariate datasets. Hence we cannot work with the raw data since the i.i.d. assumption may be inappropriate. A solution is to, instead, work on the 'innovations', which can be obtained by filtering out the volatility clustering and leverage effects from the raw return data. For each time series of market returns, we assume an exponential GARCH(1,1) model (Nelson, 1991) and fit the parameters by maximizing the quasi-likelihood corresponding to Student-t distributed innovations, denoted as z, with an unknown number of degrees of freedom. Now the Ljung-Box tests do not reject the serial independence of the original, absolute, nor squared sample innovations at the 5% level. The innovations z will also be called the filtered returns. We are interested in the *conditional*, on the information at time $t-1$, extreme quantile region of the joint raw return $\boldsymbol{r}_t = (r_t^{US}, r_t^{UK}, r_t^{JPN})$, since it describes the tail of the distribution one day ahead. This conditional quantile region can be obtained via an affine transformation from that of $\boldsymbol{z}_t = (z_t^{US}, z_t^{UK}, z_t^{JPN})$, which can be estimated directly through our approach.

Next we check the equality of the extreme value indices for the positive and negative tails of the univariate returns, implied by Assumption 2.1.1. The Hill estimates, for $k = 80$, for the right and left tails of the filtered returns in all three markets, in increasing order, are: 0.1775, 0.1779, 0.2230, 0.2247, 0.2550, 0.2614. The maximal difference is 0.0839, which, based on the asymptotic normality of the Hill estimator, corresponds to an approximate p-value of 0.165. Hence there is no evidence for the inequality of these six extreme value indices. We also test for bivariate regular variation of the filtered returns of the three possible pairs of markets using the test in Einmahl and Krajina (2016). Also these three tests do not reject the null hypothesis

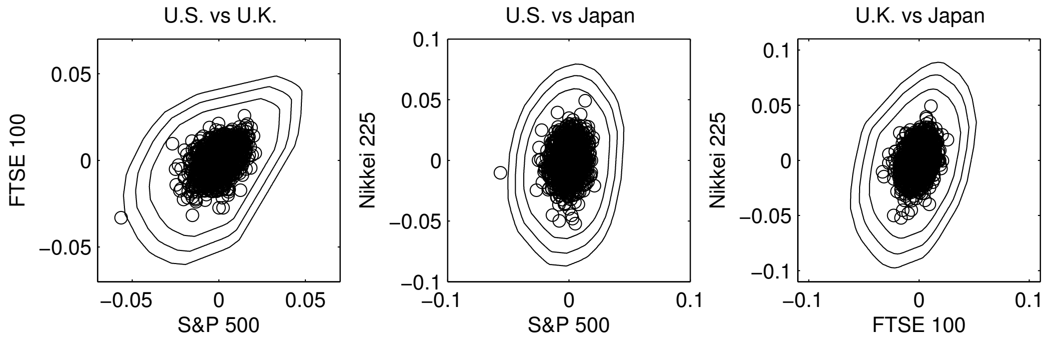


Figure 2.4: Predicted bivariate quantile regions of raw returns on July 2, 2007 (1 trading day ahead) for p=1/2000, 1/5000, 1/10000 based on the price data from July 2, 2001 to June 29, 2007. The plotted return observations are computed from the filtered returns using the predicted variance.

(at the 5% significance level).

Figure 2.4 shows the predicted bivariate extreme quantile regions/contours of raw returns for $p = 1/2000$, $1/5000$, $1/10000$ for July 2, 2007 (that is, one trading day ahead) for every pair of markets with $k = 160$. These figures convey crucial information to the risk manager. The extreme quantile regions reveal the (conditional) tail dependence structure of international capital market. Neglecting the joint behavior can lead to an overestimated diversifiability of risks across international markets and, therefore, underestimation of systematic risk; see, e.g., Longin and Solnik (2001). Furthermore, these extreme quantile regions also provide a set of unlikely scenario for stress testing (McNeil and Smith, 2012).

Another application of the depth-based extreme quantile regions is to detect data outliers. Here we consider a practical definition of outlier, namely that a data point has a rare joint innovation behavior: more precisely, its innovations lie in the (estimated) quantile region of filtered returns for an extremely small p, say 1/10000. Note that an outlier in a high dimensional space is not necessarily an outlier in its subspaces with reduced dimensions. This means the outcome depends on the choice of the data space. In our

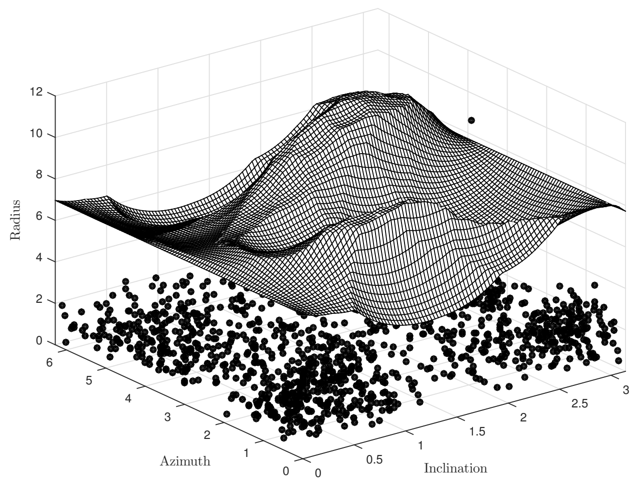


Figure 2.5: Estimated trivariate quantile regions of filtered returns with p=1/10000.

sample, we observe the biggest loss in the US market on February 27, 2007 during the well-known "Chinese Correction" event. On the same day, the Chinese market index (SSE Composite index) dropped by 9%, breaking the 10-year record. We observe from Figure 2.5 that this data point is inside the estimated, with $k = 300$, extreme trivariate quantile region for $p = 1/10000$, i.e., the space above the surface. We conclude that this point is an outlier in the three-dimensional space.

Acknowledgement We are very grateful to two referees, an Editor, and an Associate Editor for many insightful comments, questions and suggestions, which led to this greatly improved version of the manuscript.

2.5 Proofs

Proof of Proposition 1. Let $0 < p < 1$ and $\beta = \sup\{\widetilde{\beta} : P(\mathcal{Q}(\boldsymbol{X}, \widetilde{\beta})) \leq p\}$. Note that $0 < \beta < 1$. Take a sequence of positive numbers $\{\beta_m^-\}_{m=1}^{\infty}$ such that $\beta_m^- \uparrow \beta$ as $m \to \infty$. It follows that $\{\mathcal{Q}(\boldsymbol{X}, \beta_m^-)\}_{m=1}^{\infty}$ is an increasing sequence of sets. Therefore

$$P\left(\bigcup_{m=1}^{\infty} \mathcal{Q}(\boldsymbol{X}, \beta_m^-)\right) = \lim_{m\to\infty} P(\mathcal{Q}(\boldsymbol{X}, \beta_m^-)) \leq p.$$

It is easy to verify that

$$\bigcup_{m=1}^{\infty} \mathcal{Q}(\boldsymbol{X}, \beta_m^-) = \left\{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) < \beta\right\} = \mathcal{Q}(\boldsymbol{X}, \beta) \setminus \mathcal{C}_\beta.$$

Hence $P(\mathcal{Q}(\boldsymbol{X}, \beta)) = P(\mathcal{Q}(\boldsymbol{X}, \beta) \setminus \mathcal{C}_\beta) \leq p$ by Assumption 2. On the other hand, taking a sequence of numbers $\{\beta_m^+\}_{n=1}^{\infty}$ such that $\beta_m^+ \downarrow \beta$ as $m \to \infty$, analogously, it holds that

$$p \leq \lim_{m\to\infty} P(\mathcal{Q}(\boldsymbol{X}, \beta_m^+)) = P\left(\bigcap_{m=1}^{\infty} \mathcal{Q}(\boldsymbol{X}, \beta_m^+)\right) = P(\mathcal{Q}(\boldsymbol{X}, \beta)).$$

It follows that $P(\mathcal{Q}(\boldsymbol{X}, \beta)) = p$. □

We first prove Proposition 3 and then Proposition 2 and Theorem 1.

Proof of Proposition 3. (a) For boundedness and convexity we only need to examine $\widehat{S}^c$. The boundedness holds since, almost surely, $HD(\boldsymbol{u}, \widehat{\nu}^*) \leq \widehat{\nu}(H_{1,\boldsymbol{u}}) \leq 1$, $\boldsymbol{u} \in \Theta$. Next we show that $\widehat{S} = \cup_{\boldsymbol{u}\in\Theta} H_{(\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}},\boldsymbol{u}}$. Take arbitrary $\boldsymbol{x} =: r\boldsymbol{w} \in \cup_{\boldsymbol{u}\in\Theta} H_{(\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}},\boldsymbol{u}}$, with $\boldsymbol{w} \in \Theta$. Then for some $\boldsymbol{u} \in \Theta$, $\boldsymbol{x} \in H_{(\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}},\boldsymbol{u}}$, and therefore $\boldsymbol{u}^T\boldsymbol{x} = r\boldsymbol{u}^T\boldsymbol{w} \geq (\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}}$, i.e., $r \geq (\boldsymbol{u}^T\boldsymbol{w})^{-1}(\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}} = \left(\widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}})\right)^{\widehat{\gamma}} \geq (HD(\boldsymbol{w}, \widehat{\nu}^*))^{\widehat{\gamma}}$. Hence $\boldsymbol{x} \in \widehat{S}$. Now take arbitrary $\boldsymbol{x} =: r\boldsymbol{w} \in \widehat{S}$. Note that for all $\boldsymbol{w} \in \Theta$, we have $HD(\boldsymbol{w}, \widehat{\nu}^*) = \widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}})$ with some $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{w}) \in \Theta$ and it follows that $\boldsymbol{x} \in H_{(\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}},\boldsymbol{u}}$. Hence we obtain that $\widehat{S}^c = \cap_{\boldsymbol{u}\in\Theta} H^c_{(\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}},\boldsymbol{u}}$ is convex.

(b) It suffices to prove the orthogonal and scale equivariance separately. The orthogonal transformation has no impact on the radii $R_1, \ldots, R_n$ of the random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. It is easy to verify that the only change is

$\widehat{S}_{\boldsymbol{RX}} = \boldsymbol{R}\widehat{S}_{\boldsymbol{X}}$, then the orthogonal equivariance follows. The scale equivariance comes in a similar way by using the facts $\widehat{U}_{c\boldsymbol{X}}(n/k) = c\widehat{U}_{\boldsymbol{X}}(n/k)$ and other components of the estimate remain the same.

(c) Straightforward.

□

To prove Proposition 2 we need some lemmas for which we assume that the conditions of the proposition hold.

Lemma 2.5.1.

$$\lim_{t\to\infty}\sup_{\boldsymbol{u}\in\Theta}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH_{1,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)} - \nu(H_{1,\boldsymbol{u}})\right| = 0.$$

Proof. Lemma 1 in Einmahl et al. (2015b) yields

$$\lim_{t\to\infty}\sup_{\boldsymbol{u}\in\Theta}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH_{1,\boldsymbol{u}})}{ct^{-1/\gamma}} - \nu(H_{1,\boldsymbol{u}})\right| = 0.$$

Now the result follows from Assumption 3. □

Lemma 2.5.2. *For all $\varepsilon > 0$,*

$$\lim_{t\to\infty}\sup_{H\in\mathcal{H}^\varepsilon}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH)}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)} - \nu(H)\right| = 0,$$

where $\mathcal{H}^\varepsilon = \{H_{r,\boldsymbol{u}} \in \mathcal{H} : r \geq \varepsilon\}$.

Proof. For $r \geq \varepsilon > 0$,

$$\begin{aligned}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH_{r,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)} - \nu(H_{r,\boldsymbol{u}})\right| \leq & \frac{\mathbb{P}(\boldsymbol{X}\in trH_{1,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq tr)}\left|\frac{\mathbb{P}(\|\boldsymbol{X}\|\geq tr)}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)} - r^{-1/\gamma}\right| \\ & + r^{-1/\gamma}\left|\frac{\mathbb{P}(\boldsymbol{X}\in trH_{1,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq tr)} - \nu(H_{1,\boldsymbol{u}})\right|.\end{aligned}$$

The result follows from Lemma 2.5.1 [cf. Theorem 2.1 in de Haan and Resnick (1987)]. □

Lemma 2.5.3. *The function $\nu(H_{1,\cdot})$ is continuous on Θ and hence $\delta_0 := \inf_{\boldsymbol{u}\in\Theta}\nu(H_{1,\boldsymbol{u}}) > 0$.*

Proof. Take arbitrary $\boldsymbol{u}$, $\boldsymbol{v} \in \Theta$ such that $\delta := \|\boldsymbol{u}-\boldsymbol{v}\| \in (0,1)$. Note that

$$H_{1,\boldsymbol{u}} \setminus H_{1-\delta^{1/2},\boldsymbol{v}} \subset \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{u}-\boldsymbol{v})^T \boldsymbol{x} \geq \delta^{1/2}\} \subset \delta^{-1/2}\mathbb{C}$$

where $\mathbb{C} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \geq 1\}$. It follows that

$$\begin{aligned}
\nu(H_{1,\boldsymbol{u}} \setminus H_{1,\boldsymbol{v}}) &\leq \nu(H_{1,\boldsymbol{u}} \setminus H_{1-\delta^{1/2},\boldsymbol{v}}) + \nu(H_{1-\delta^{1/2},\boldsymbol{v}} \setminus H_{1,\boldsymbol{v}}) \\
&\leq \nu(\delta^{-1/2}\mathbb{C}) + [(1-\delta^{1/2})^{-1/\gamma} - 1]\nu(H_{1,\boldsymbol{v}}) \\
&\leq \delta^{1/(2\gamma)} + [(1-\delta^{1/2})^{-1/\gamma} - 1]
\end{aligned}$$

and, analogously, $\nu(H_{1,\boldsymbol{v}} \setminus H_{1,\boldsymbol{u}})$ can be bounded by the same number. Hence the continuity follows since $|\nu(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{v}})| \leq \nu(H_{1,\boldsymbol{u}} \setminus H_{1,\boldsymbol{v}}) + \nu(H_{1,\boldsymbol{v}} \setminus H_{1,\boldsymbol{u}})$ can be made arbitrarily small for sufficiently small δ. The continuity of $\nu(H_{1,\cdot})$ on the compact Θ in combination with the last part of Assumption 1 yields $\delta_0 > 0$. $\square$

Lemma 2.5.4. *Let δ be a constant such that $0 < \delta \leq \delta_0^\gamma$ and let $\varepsilon > 0$. Then for all $\boldsymbol{z} \in \mathbb{R}^d$ with $\|\boldsymbol{z}\| \geq \varepsilon$,*

$$HD(\boldsymbol{z}, \nu) = \inf_{\boldsymbol{u}^T\boldsymbol{z} \geq \delta\varepsilon} \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})$$

and there exists a $M > 0$, which only depends on $\delta\varepsilon$, such that for all $t \geq M$

$$HD(t\boldsymbol{z}, P) = \inf_{\boldsymbol{u}^T\boldsymbol{z} \geq \delta\varepsilon} P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}).$$

Proof. We only prove the second part. The proof of the first part is similar. For $\|\boldsymbol{z}\| \geq \varepsilon$, by Lemma 2.5.2 we have

$$\inf_{\boldsymbol{u}^T\boldsymbol{z} < \delta\varepsilon} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} \geq \inf_{\boldsymbol{u} \in \Theta} \left\{ \frac{P(H_{t\delta\varepsilon,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} \to \inf_{\boldsymbol{u} \in \Theta} \nu\left(H_{\delta\varepsilon,\boldsymbol{u}}\right) = (\delta\varepsilon)^{-1/\gamma}\delta_0 \geq \varepsilon^{-1/\gamma}$$

and

$$\begin{aligned}
\inf_{\boldsymbol{u} \in \Theta} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} \leq \frac{P(H_{t\|\boldsymbol{z}\|,\boldsymbol{z}/\|\boldsymbol{z}\|})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \to \nu(H_{\|\boldsymbol{z}\|,\boldsymbol{z}/\|\boldsymbol{z}\|}) &= \|\boldsymbol{z}\|^{-1/\gamma}\, \nu(H_{1,\boldsymbol{z}/\|\boldsymbol{z}\|}) \\
&\leq \varepsilon^{-1/\gamma}(1-\delta_0),
\end{aligned}$$

where in last step we use the fact $\nu(H_{1,\boldsymbol{w}}) \le \nu(\mathbb{B}^c) - \nu(H_{1,-\boldsymbol{w}}) \le 1 - \delta_0$. It then follows from Lemma 2.5.2 that there exists a $M = M_{\delta\varepsilon} > 0$, such that for all $t \ge M$

$$\inf_{\boldsymbol{u}^T\boldsymbol{z}<\delta\varepsilon} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},u})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \right\} > \varepsilon^{-1/\gamma}\left(1 - \frac{\delta_0}{2}\right) > \inf_{\boldsymbol{u}\in\Theta} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \right\}.$$

This implies that $\inf_{\boldsymbol{u}^T\boldsymbol{z}<\delta\varepsilon} P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) > \inf_{\boldsymbol{u}\in\Theta} P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) = HD(t\boldsymbol{z}, P)$ and consequently the second part of the lemma. □

Proof of Proposition 2. From Lemma 2.5.4 with $\delta = \delta_0^\gamma$, we know that for sufficiently large t

$$\begin{aligned}
&\sup_{\|\boldsymbol{z}\|\ge\varepsilon} \left| \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} - HD(\boldsymbol{z}, \nu) \right| \\
&= \sup_{\|\boldsymbol{z}\|\ge\varepsilon} \left| \inf_{\boldsymbol{u}^T\boldsymbol{z}\ge\delta_0^\gamma\varepsilon} \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} - \inf_{\boldsymbol{u}^T\boldsymbol{z}\ge\delta_0^\gamma\varepsilon} \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) \right| \\
&\le \sup_{\|\boldsymbol{z}\|\ge\varepsilon} \sup_{\boldsymbol{u}^T\boldsymbol{z}\ge\delta_0^\gamma\varepsilon} \left| \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} - \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) \right|.
\end{aligned}$$

The rest follows from Lemma 2.5.2. □

To prove Theorem 1 we need some further lemmas. In the sequel we will always assume that the conditions of the theorem hold.

Lemma 2.5.5. *For each $\varepsilon > 0$, there exists $t_0 > 0$ such that for $t > t_0$*

$$\left\{ \boldsymbol{z} \in \mathbb{R}^d : \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \le \varepsilon \right\} \subset \left\{ \boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| > \left(\frac{\delta_0}{2\varepsilon}\right)^\gamma \right\}.$$

Proof. It suffices to prove, for large t

$$\left\{ \boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| \le \left(\frac{\delta_0}{2\varepsilon}\right)^\gamma \right\} \subset \left\{ \boldsymbol{z} \in \mathbb{R}^d : \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} > \varepsilon \right\}.$$

Write $\delta = (\delta_0/2\varepsilon)^\gamma$. Take any $\boldsymbol{z} \in \mathbb{R}^d$ with $\|\boldsymbol{z}\| \le \delta$. Lemma 2.5.2 yields

$$\inf_{\boldsymbol{u}\in\Theta} \frac{P(tH_{\delta,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \to \inf_{\boldsymbol{u}\in\Theta} \nu(H_{\delta,\boldsymbol{u}}) = \delta^{-1/\gamma} \inf_{\boldsymbol{u}\in\Theta} \nu(H_{1,\boldsymbol{u}}) = \delta^{-1/\gamma}\delta_0 = 2\varepsilon.$$

Hence there exists a $t_0 > 0$ such that for $t > t_0$

$$\frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} = \inf_{\boldsymbol{u}\in\Theta} \frac{P(tH_{\boldsymbol{u}^T z,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \ge \inf_{\boldsymbol{u}\in\Theta} \frac{P(tH_{\delta,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} > 2\varepsilon - \varepsilon = \varepsilon.$$

□

Lemma 2.5.6. *As $n \to \infty$, $p/\beta \to \nu(S)$.*

Proof. Under Assumption 1, $\mathbb{P}\left(\|\boldsymbol{X}\| \geq U\left(1/\beta\right)\right)/\beta \to 1$, as $n \to \infty$. Hence, as $n \to \infty$,

$$\begin{aligned} p =& P(\mathcal{Q}) = P(\{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) \leq \beta\}) \\ =& P(\{U(1/\beta)\boldsymbol{z} \in \mathbb{R}^d : HD(U(1/\beta)\boldsymbol{z}, P) \leq \beta\}) \\ \sim& P\left(U(1/\beta)\left\{\boldsymbol{z} \in \mathbb{R}^d : \frac{HD(U(1/\beta)\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \geq U(1/\beta))} \leq 1\right\}\right). \end{aligned}$$

By Lemma 2.5.5 we know, when n is large,

$$S_n := \left\{\boldsymbol{z} \in \mathbb{R}^d : \frac{HD(U(1/\beta)\boldsymbol{z}; P)}{\mathbb{P}(\|\boldsymbol{X}\| \geq U(1/\beta))} \leq 1\right\} \subset \left\{\boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| > \left(\frac{\delta_0}{2}\right)^{\gamma}\right\}.$$

Then Proposition 2 yields that for any $\varepsilon > 0$ there exists an M_ε such that when $n > M_\varepsilon$,

$$(1+\varepsilon)S \subset S_n \subset (1-\varepsilon)S.$$

Therefore

$$\lim_{n\to\infty}\left|\frac{p}{\beta} - \nu(S)\right| \leq \nu((1-\varepsilon)S) - \nu((1+\varepsilon)S) = \nu(S)((1-\varepsilon)^{-1/\gamma} - (1+\varepsilon)^{-1/\gamma}),$$

which immediately implies our result since ε is arbitrary. $\square$

Lemma 2.5.7. *As $n \to \infty$,*

$$\sup_{\boldsymbol{w}\in\Theta}\left|\frac{HD(\boldsymbol{w}, \widehat{\nu}^*)}{HD(\boldsymbol{w}, \nu)} - 1\right| \xrightarrow{\mathbb{P}} 0.$$

Proof. First we show, as $n \to \infty$,

$$\sup_{\boldsymbol{u}\in\Theta}|\widehat{\nu}(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{u}})| \xrightarrow{\mathbb{P}} 0. \tag{2.5.1}$$

Note that

$$\begin{aligned} &\sup_{\boldsymbol{u}\in\Theta}|\widehat{\nu}(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{u}})| \\ \leq& \sup_{\boldsymbol{u}\in\Theta}\left|\frac{n}{k}P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right) - \frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)\right| \end{aligned}$$

$$
\begin{aligned}
&+\sup_{\boldsymbol{u}\in\Theta}\left|\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)-\nu(H_{1,\boldsymbol{u}})\right| \\
\leq&\sup_{\boldsymbol{u}\in\Theta}\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right| \\
&+\sup_{\boldsymbol{u}\in\Theta}\left|\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)-\nu(H_{1,u})\right|.
\end{aligned}
$$

From Lemma 2.5.1 we know for (2.5.1) it suffices to show

$$
\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|=\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}\frac{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|\xrightarrow{\mathbb{P}}0.
$$

In other words, it suffices to show

$$
\mathrm{I}:=\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|\xrightarrow{\mathbb{P}}0 \text{ and } \mathrm{II}:=\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|\xrightarrow{\mathbb{P}}0.
$$

For any $0<\eta<1$, define events $\Omega_n=\{(1-\eta)U\left(\frac{n}{k}\right)\leq\widehat{U}\left(\frac{n}{k}\right)\leq(1+\eta)U\left(\frac{n}{k}\right)\}$. Then it follows that $\mathbb{P}(\Omega_n)\to 1$ as $n\to\infty$, since $\widehat{U}\left(\frac{n}{k}\right)/U\left(\frac{n}{k}\right)\xrightarrow{\mathbb{P}}1$. On Ω_n, we have

$$
(1+\eta)U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\subset\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\subset(1-\eta)U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}.
$$

Denoting $\mathcal{H}_{1+\eta}=\{H_{r,\boldsymbol{u}}\in\mathcal{H}:r\leq 1+\eta,\boldsymbol{u}\in\Theta\}$, we have

$$
\inf_{H\in\mathcal{H}_{1+\eta}}\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H\right)\to\inf_{H\in\mathcal{H}_{1+\eta}}\nu(H)=(1+\eta)^{-1/\gamma}\inf_{\boldsymbol{u}\in\Theta}\nu(H_{1,\boldsymbol{u}})=:2\delta.
$$

Note that $\mathcal{H}_{1+\eta}$ is a VC class and that condition (5.1) in Alexander (1987) holds with $\gamma_n=k\delta/n$ since it can be shown that $g(\gamma_n)$ is bounded, as $n\to\infty$. Applying Theorem 5.1 of that paper yields that

$$
\mathrm{I}\leq\sup_{H\in\mathcal{H}_{1+\eta}}\left|\frac{P_n\left(U\left(\frac{n}{k}\right)H\right)}{P\left(U\left(\frac{n}{k}\right)H\right)}-1\right|\leq\sup_{P(H)\geq\frac{k\delta}{n}}\left|\frac{P_n(H)}{P(H)}-1\right|\xrightarrow{\mathbb{P}}0.
$$

On the other hand, on Ω_n, for any $\boldsymbol{u}\in\Theta$

$$
\frac{P\left(U\left(\frac{n}{k}\right)H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}\leq\frac{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}\leq\frac{P\left(U\left(\frac{n}{k}\right)H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}.
$$

Then

$$\begin{aligned}
\mathrm{II} &\leq \sup_{\boldsymbol{u}\in\Theta}\left\{\max\left\{1-\frac{P\left(U\left(\frac{n}{k}\right)H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}, \frac{P\left(U\left(\frac{n}{k}\right)H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right\}\right\} \\
&\leq \max\left\{1-\inf_{\boldsymbol{u}\in\Theta}\frac{P\left(U\left(\frac{n}{k}\right)H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}, \sup_{\boldsymbol{u}\in\Theta}\frac{P\left(U\left(\frac{n}{k}\right)H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right\} \\
&=: \max\{\mathrm{II}_1, \mathrm{II}_2\}.
\end{aligned}$$

Note that

$$\inf_{\boldsymbol{u}\in\Theta}\frac{P\left(U\left(\frac{n}{k}\right)H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)} = \inf_{\boldsymbol{u}\in\Theta}\frac{\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1+\eta,\boldsymbol{u}}\right)}{\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)} \to \inf_{\boldsymbol{u}\in\Theta}\frac{\nu(H_{1+\eta,\boldsymbol{u}})}{\nu(H_{1,\boldsymbol{u}})} = (1+\eta)^{-1/\gamma}$$

and similarly

$$\sup_{\boldsymbol{u}\in\Theta}\frac{P\left(U\left(\frac{n}{k}\right)H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)} \to (1-\eta)^{-1/\gamma}.$$

Since η can be arbitrarily small, it follows from above that both II_1 and II_2 can be arbitrarily small when n is sufficiently large. This implies that $\mathrm{II} \xrightarrow{\mathbb{P}} 0$. Hence (2.5.1) holds. For the rest now it is sufficient to show

$$\sup_{\boldsymbol{w}\in\Theta}|HD(\boldsymbol{w},\widehat{\nu}^*) - HD(\boldsymbol{w},\nu)| \xrightarrow{\mathbb{P}} 0. \tag{2.5.2}$$

Note that, if (2.5.2) is true, we are done since

$$\inf_{\boldsymbol{w}\in\Theta} HD(\boldsymbol{w},\nu) = \inf_{\boldsymbol{w}\in\Theta}\inf_{\boldsymbol{u}\in\Theta}\{\nu(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}})\} \geq \inf_{\boldsymbol{u}\in\Theta}\{\nu(H_{1,\boldsymbol{u}})\} > 0.$$

Take a $\delta > 0$ such that $0 < \delta < \delta_0^\gamma$. From Lemma 2.5.4 we know for $\boldsymbol{w} \in \Theta$

$$HD(\boldsymbol{w},\nu) = \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta}\{\nu(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}})\}.$$

Next, define the events $\widetilde{\Omega}_n = \{(\inf_{\boldsymbol{u}\in\Theta}\widehat{\nu}(H_{1,\boldsymbol{u}}))^{\widehat{\gamma}} > \delta\}$. It holds that $\mathbb{P}(\widetilde{\Omega}_n) \to 1$, $n \to \infty$, by the uniform convergence of $\widehat{\nu}$ from (2.5.1) and the consistency of $\widehat{\gamma}$. Analogously to Lemma 2.5.4 we also have, on $\widetilde{\Omega}_n$,

$$HD(\boldsymbol{w},\widehat{\nu}^*) = \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta}\{\widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}})\}.$$

Then, noting that $\boldsymbol{u}^T\boldsymbol{w} \leq 1$ for $\boldsymbol{u}, \boldsymbol{w} \in \Theta$, on $\widetilde{\Omega}_n$

$$
\begin{aligned}
&\sup_{\boldsymbol{w}\in\Theta} |HD(\boldsymbol{w}, \widehat{\nu}^*) - HD(\boldsymbol{w}, \nu)| \\
&= \sup_{\boldsymbol{w}\in\Theta} \left| \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) - \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \nu(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) \right| \\
&\leq \sup_{\boldsymbol{w}\in\Theta} \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \left| \widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) - \nu(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) \right| \\
&= \sup_{\boldsymbol{w}\in\Theta} \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \left| (\boldsymbol{u}^T\boldsymbol{w})^{-1/\widehat{\gamma}} \widehat{\nu}(H_{1,\boldsymbol{u}}) - (\boldsymbol{u}^T\boldsymbol{w})^{-1/\gamma} \nu(H_{1,\boldsymbol{u}}) \right| \\
&\leq \sup_{\boldsymbol{u}\in\Theta} \widehat{\nu}(H_{1,\boldsymbol{u}}) \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} (\boldsymbol{u}^T\boldsymbol{w})^{-1/\gamma} \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \left| (\boldsymbol{u}^T\boldsymbol{w})^{1/\gamma - 1/\widehat{\gamma}} - 1 \right| \\
&\quad + \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} (\boldsymbol{u}^T\boldsymbol{w})^{-1/\gamma} \sup_{\boldsymbol{u}\in\Theta} |\widehat{\nu}(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{u}})| \\
&\leq \delta^{-1/\gamma} \left| \delta^{1/\gamma - 1/\widehat{\gamma}} - 1 \right| + \delta^{-1/\gamma} \sup_{\boldsymbol{u}\in\Theta} |\widehat{\nu}(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{u}})|.
\end{aligned}
$$

By the consistency of $\widehat{\gamma}$ and (2.5.1) we can conclude that (2.5.2) is true. $\square$

Lemma 2.5.8. *As $n \to \infty$,*

$$
\widehat{\nu}(\widehat{S}) \xrightarrow{\mathbb{P}} \nu(S).
$$

Proof. Let $\varepsilon > 0$. Applying Chebyshev's inequality yields, for any $\delta > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\left| \frac{n}{k} P_n \left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) - \frac{n}{k} P\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) \right| \geq \delta \right) \\
&= \mathbb{P}\left(\left| P_n \left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) - P\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) \right| \geq \frac{k}{n}\delta \right) \\
&\leq \frac{P\left(U(n/k)(1+\varepsilon)^2 S\right)\left[1 - P\left(U(n/k)(1+\varepsilon)^2 S\right)\right]}{n(k\delta/n)^2} \\
&= \frac{1}{k} \frac{P\left(U(n/k)(1+\varepsilon)^2 S\right)}{\mathbb{P}\left(\|\boldsymbol{X}\| \geq U(n/k)\right)} [1 - P\left(U(n/k)(1+\varepsilon)^2 S\right)] \frac{1}{\delta^2} \to 0
\end{aligned}
$$

and therefore

$$
\frac{n}{k} P_n \left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) \xrightarrow{\mathbb{P}} \nu((1+\varepsilon)^2 S) = (1+\varepsilon)^{-2/\gamma} \nu(S).
$$

Similarly,

$$
\frac{n}{k} P_n \left(U\left(\frac{n}{k}\right)(1-\varepsilon)^2 S \right) \xrightarrow{\mathbb{P}} \nu((1-\varepsilon)^2 S) = (1-\varepsilon)^{-2/\gamma} \nu(S).
$$

Define events

$$\Omega_n = \{(1+\varepsilon)^2 U(n/k)S \subset \widehat{U}(n/k)\widehat{S} \subset (1-\varepsilon)^2 U(n/k)S\}$$

then $P(\Omega_n) \to 1$ because of $\widehat{U}(n/k)/U(n/k) \xrightarrow{\mathbb{P}} 1$ and Lemma 2.5.7. On Ω_n,

$$\begin{aligned}
\left|\widehat{\nu}(\widehat{S}) - \nu(S)\right| &\le \left|\frac{n}{k}P_n(\widehat{U}(n/k)\widehat{S}) - \frac{n}{k}P(U(n/k)S)\right| + \left|\frac{n}{k}P(U(n/k)S) - \nu(S)\right| \\
&\le \left|\frac{n}{k}P_n((1-\varepsilon)^2 U(n/k)S) - \frac{n}{k}P(U(n/k)S)\right| \\
&\quad + \left|\frac{n}{k}P(U(n/k)S) - \frac{n}{k}P_n((1+\varepsilon)^2 U(n/k)S)\right| \\
&\quad + \left|\frac{n}{k}P(U(n/k)S) - \nu(S)\right| \\
&\xrightarrow{\mathbb{P}} [(1-\varepsilon)^{-2/\gamma} - (1+\varepsilon)^{-2/\gamma}]\nu(S)
\end{aligned}$$

where ε can be chosen arbitrarily small. Hence $\widehat{\nu}(\widehat{S}) \xrightarrow{\mathbb{P}} \nu(S)$. □

Proof of Theorem 1. Define

$$\widehat{r}_n^{\boldsymbol{w}} := \frac{\widehat{U}(n/k)(k\widehat{\nu}(\widehat{S})/(np))^{\widehat{\gamma}}}{U(1/\beta)}(HD(\boldsymbol{w}, \widehat{\nu}^*))^{\widehat{\gamma}}.$$

Note that, as $n \to \infty$, the continuity of U yields $\widehat{U}(n/k)/U(n/k) \xrightarrow{\mathbb{P}} 1$ while Lemma 2.5.8 and the consistency of $\widehat{\gamma}$ imply that $(k\widehat{\nu}(\widehat{S})/np)^{\widehat{\gamma}}/(k\nu(S)/np)^{\gamma} \xrightarrow{\mathbb{P}} 1$. Moreover, Assumption 3 gives that, as $n \to \infty$,

$$U(n/k)(k/n)^{\gamma} \to c^{\gamma} \quad \text{and} \quad U(1/\beta)\beta^{\gamma} \to c^{\gamma}.$$

Hence, by Lemmas 2.5.6 and 2.5.8, it holds that

$$\begin{aligned}
\frac{\widehat{U}(n/k)(k\widehat{\nu}(\widehat{S})/np)^{\widehat{\gamma}}}{U(1/\beta)} &= \frac{\widehat{U}(n/k)(k\widehat{\nu}(\widehat{S})/np)^{\widehat{\gamma}}}{U(n/k)(k\nu(S)/np)^{\gamma}} \frac{U(n/k)(k/n)^{\gamma}}{U(1/\beta)\beta^{\gamma}} \left(\frac{\beta\nu(S)}{p}\right)^{\gamma} \\
&\xrightarrow{\mathbb{P}} 1 \cdot \frac{c^{\gamma}}{c^{\gamma}} \cdot 1^{\gamma} = 1.
\end{aligned}$$

Combining this with Lemma 2.5.7 and writing $r^{\boldsymbol{w}} = (HD(\boldsymbol{w}, \nu))^{\gamma}$, we obtain

$$\sup_{\boldsymbol{w} \in \Theta} \left|\frac{\widehat{r}_n^{\boldsymbol{w}}}{r^{\boldsymbol{w}}} - 1\right| \xrightarrow{\mathbb{P}} 0.$$

This implies that, for any $\varepsilon > 0$, the probability of the events $\Omega_n = \{(1-\varepsilon)r^{\boldsymbol{w}} \le \widehat{r}_n^{\boldsymbol{w}} \le (1+\varepsilon)r^{\boldsymbol{w}}, \text{ for all } \boldsymbol{w} \in \Theta\}$ converges to 1 as $n \to \infty$. Then, on Ω_n for large n,

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} \left| \frac{HD(\boldsymbol{x},P)}{\beta} - 1 \right| \\
&= \sup_{\boldsymbol{w}\in\Theta} \left| \frac{HD(U(1/\beta)\widehat{r}_n^{\boldsymbol{w}}\boldsymbol{w},P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge U(1/\beta))} - HD(HD(\boldsymbol{w},\nu)^{\gamma}\boldsymbol{w},\nu) \right| (1+o(1)) + o(1) \\
&\le \sup_{\boldsymbol{w}\in\Theta} \left| \frac{HD(U(1/\beta)r^{\boldsymbol{w}}\boldsymbol{w},P)}{P(\|\boldsymbol{X}\| \ge U(1/\beta))} - HD(r^{\boldsymbol{w}}\boldsymbol{w},\nu) \right| (1+o(1)) \\
&\quad + \sup_{\boldsymbol{w}\in\Theta} \left| \frac{HD((1-\varepsilon)U(1/\beta)r^{\boldsymbol{w}}\boldsymbol{w},P)}{P(\|\boldsymbol{X}\| \ge U(1/\beta))} - \frac{HD((1+\varepsilon)U(1/\beta)r^{\boldsymbol{w}}\boldsymbol{w},P)}{P(\|\boldsymbol{X}\| \ge U(1/\beta))} \right| (1+o(1)) \\
&\quad + o(1) \\
&\quad =: \mathrm{I} + \mathrm{II} + o(1).
\end{aligned}
$$

Here the $o(1)$-terms stem from the convergence $\mathbb{P}(\|\boldsymbol{X}\| \ge U(1/\beta))/\beta \to 1$ $(n \to \infty)$. By Proposition 2, we know $\mathrm{I} \xrightarrow{\mathbb{P}} 0$ and

$$
\mathrm{II} \xrightarrow{\mathbb{P}} [(1-\varepsilon)^{-1/\gamma} - (1+\varepsilon)^{-1/\gamma}]HD(r^{\boldsymbol{w}}\boldsymbol{w},\nu) = (1-\varepsilon)^{-1/\gamma} - (1+\varepsilon)^{-1/\gamma}.
$$

Since ε can be arbitrarily small, it holds that, as $n \to \infty$,

$$
\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} \left| \frac{HD(\boldsymbol{x},P)}{\beta} - 1 \right| \xrightarrow{\mathbb{P}} 0,
$$

which immediately implies the first part of the theorem.

Next, we show that the second part of the theorem follows from the first part. Write $\widehat{\beta}_n^{+} = \sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} HD(\boldsymbol{x},P)$ and $\widehat{\beta}_n^{-} = \inf_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} HD(\boldsymbol{x},P)$. Because of the nestedness of $\mathcal{Q}(\boldsymbol{X},\beta)$, by Theorem 2.11 in Zuo and Serfling (2000a), we have $\mathcal{Q}(\boldsymbol{X},\widehat{\beta}_n^{-}) \subset \widehat{\mathcal{Q}}_n \subset \mathcal{Q}(\boldsymbol{X},\widehat{\beta}_n^{+})$ surely. Again using this nestedness we have, surely,

$$
\begin{aligned}
\frac{P(\widehat{\mathcal{Q}}_n\Delta\mathcal{Q})}{p} &\le \frac{P(\mathcal{Q}\Delta\mathcal{Q}(\boldsymbol{X},\widehat{\beta}_n^{+}))}{p} + \frac{P(\mathcal{Q}\Delta\mathcal{Q}(\boldsymbol{X},\widehat{\beta}_n^{-}))}{p} \\
&= \left| \frac{P\mathcal{Q}(\boldsymbol{X},\widehat{\beta}_n^{+}) - P\mathcal{Q}}{p} \right| + \left| \frac{P\mathcal{Q}(\boldsymbol{X},\widehat{\beta}_n^{-}) - P\mathcal{Q}}{p} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \frac{P\mathcal{Q}(\boldsymbol{X}, \widehat{\beta}_n^+)}{\widehat{\beta}_n^+} \frac{\widehat{\beta}_n^+}{\beta} \frac{\beta}{p} - 1 \right| + \left| \frac{P\mathcal{Q}(\boldsymbol{X}, \widehat{\beta}_n^-)}{\widehat{\beta}_n^-} \frac{\widehat{\beta}_n^-}{\beta} \frac{\beta}{p} - 1 \right| \\
&=: I + II.
\end{aligned}
$$

The first part of the theorem implies that $\widehat{\beta}_n^+/\beta \xrightarrow{\mathbb{P}} 1$ as $n \to \infty$. It follows that $\widehat{\beta}_n^+ \xrightarrow{\mathbb{P}} 0$ and then, similar to Lemma 2.5.6, $P\mathcal{Q}(\boldsymbol{X}, \widehat{\beta}_n^+)/\widehat{\beta}_n^+ \xrightarrow{\mathbb{P}} \nu(S)$ as $n \to \infty$. Hence, together with Lemma 2.5.6, we have $I \xrightarrow{\mathbb{P}} \left|\nu(S) \cdot 1 \cdot \frac{1}{\nu(S)} - 1\right|$, as $n \to \infty$. Similarly, $II \xrightarrow{\mathbb{P}} 0$, $n \to \infty$. $\square$

Chapter 3

Asymptotics for Extreme Depth-based Quantile Region Estimation

Abstract. Consider the small-probability quantile region in arbitrary dimensions consisting of extremely outlying points with nearly zero data depth value. Since its estimation involves extrapolation outside the data cloud, an entirely nonparametric method often fails. Using extreme value statistics, we extend the semiparametric estimation procedures in Cai et al. (2011) and He and Einmahl (2016) to incorporate various depth functions. Under weak regular variation conditions, a general consistency result is derived. To construct confidence sets that asymptotically cover the extreme quantile region or/and its complement with a pre-specified probability, we introduce new notions of distance between our estimated and true quantile region and prove their asymptotic normality via an approximation using the extreme value index only. Refined asymptotics are derived particularly for the half-space depth to include the shape estimation uncer-

tainty. The finite-sample coverage probabilities of our asymptotic confidence sets are evaluated in a simulation study for the half-space depth and the projection depth. We also apply our method to financial data.

3.1 Introduction

Associated with a probability distribution P on $\mathbb{R}^d$ $(d \geq 1)$, a data depth is a P-based function from $\mathbb{R}^d$ to $[0, \infty)$, denoted as $D(\cdot) = D(\cdot; P)$, such that provides a *center-outward ordering* in $\mathbb{R}^d$. This interpretation suggests that a relevant 'center' (also called median) with maximal depth value is available, and low/high depth corresponds to outlyingness/centrality relative to that center. For more discussions of its general notions we refer to Liu et al. (1999), Zuo and Serfling (2000a), Serfling (2006) and the many references therein.

Consider the *extreme* depth-based *quantile region* consisting of the extremely outlying points, that is, of the form

$$\mathcal{Q} = \mathcal{Q}(\mathbf{X}, \beta) = \{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}) < \beta\}$$

for a given, small probability $p = P\mathcal{Q} > 0$. Any particular choice of the depth function leads to a specific class of quantile region and the depth value $\beta = \beta(p)$ remains implicit in general. Introduced in Liu et al. (1999), the complement of $\mathcal{Q}$ is the $(1-p)$th *central region* given by

$$\mathcal{Q}^c = \{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}) \geq \beta\},$$

which enjoys many desirable equivariance and structural properties such as convexity and boundedness, under suitable regularity conditions; see, e.g., Zuo and Serfling (2000b).

For simplicity, we assume the existence of $\mathcal{Q}$ (and $\mathcal{Q}^c$) throughout the paper.

Assumption 3.1.1. There exists a $\beta = \beta(p) > 0$ such that $P\mathcal{Q}(\mathbf{X}, \beta) = p$ for all $p \in (0, 1)$.

A sufficient condition for Assumption 3.1.1 to hold (see Proposition 2.2.1) is that, for all $\beta \in (0, \infty)$,

$$P(\{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}) = \beta\}) = 0,$$

which is a crucial requirement in the uniform depth contour convergence theorem in He and Wang (1997).

It is the purpose of this paper to establish an estimation and inference procedure for $\mathcal{Q}$ based on n i.i.d. copies of $\mathbf{X}$. In the spirit of extreme value statistics, we consider p and β both to be very small in the sense that, as the sample size $n \to \infty$, $p = p_n \to 0$ ($\beta = \beta_n \to 0$) and typically at an order of $1/n$. It means that $\mathcal{Q} = \mathcal{Q}_n$ depends on n, and it contains little or even no data points. We extend the semiparametric estimation procedures proposed in Cai et al. (2011) and Chapter 2 to incorporate various depth functions and obtain a general consistency result under weak regular variation conditions.

We also provide several asymptotic results for constructing (conservative) confidence sets that asymptotically cover the quantile region $\mathcal{Q}$ or the central region $\mathcal{Q}^c$, or both simultaneously, with (at least) a prespecified probability under weak regular variation conditions. A general approach is to construct the what-we-called *naive* asymptotic confidence sets based on the asymptotic normality of the what-we-called *(directed) logarithmic distance* between our estimated and true quantile region via an approximation using the extreme value index only. In our simulation study, the finite-sample coverage probabilities of these *naive* asymptotic confidence sets are reasonably close to the correct levels for the *projection depth* but not for the *half-space depth* (the definition of these two depth functions will follow soon). The actual estimation error is substantially underestimated for the half-space depth based quantile region, whose shape is determined mostly by the *tail* behavior of the underlying distribution that is not well estimated (see Chapter 2 for more discussions). This motivates a *refined* asymptotic theory with an adjusted estimator of $\mathcal{Q}$ for the half-space depth. Specifically, we first construct some (simultaneous) asymptotically *conservative* confidence sets of $\mathcal{Q}$ or/and $\mathcal{Q}^c$, and then investigate the additional conditions under which they are also asymptotically *correct*. In the latter case we will refer to these sets as *refined* asymptotic confidence sets.

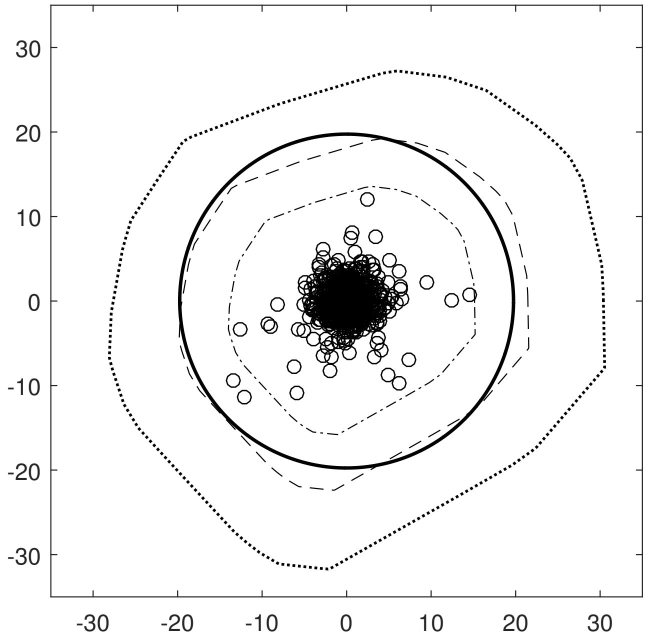


Figure 3.1: True (solid) and estimated (dashed) half-space depth based quantile regions, and simultaneous 75% refined asymptotic confidence sets of the quantile region (open outer region of dashed-dotted line) and the central region (closed inner region of the dotted line) at level $p = 1/n$ for a bivariate Student t_3 random sample; $n = 1500$.

Figure 3.1 presents an example of our extreme estimate of the half-space depth based quantile region at level $p = 1/n$ for a bivariate Student t_3 random sample with $n = 1500$. The data scarcity in the corresponding quantile region, i.e. the outer region of the circle, is clearly demonstrated. The open outer region of the dashed-dotted line is a refined asymptotic confidence set of $\mathcal{Q}$ and the closed inner region of the dotted line is the one for $\mathcal{Q}^c$. Involving extrapolation outside the data cloud, $\mathcal{Q}$ can hardly be estimated via a fully nonparametric approach. Our extreme value method can be viewed as a semiparametric approach based on some smoothed version of the empirical probability measure that is supported on the whole $\mathbb{R}^d$.

We consider multivariate regularly varying depth functions since our interest is in extreme quantile regions that are far away from the distribution center and the origin. Denote $\mathbf{0} = (0, \ldots, 0)$ as the zero vector, and $\Theta = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ as the unit sphere of the usual Euclidean norm $\|\cdot\|$.

Assumption D. For some function h: $(0, \infty) \to (0, \infty)$ regularly varying at infinity with index $-1/\xi < 0$ (Definition B.1.1 in de Haan and Ferreira, 2006), there exists a function $w : \mathbb{R}^d \setminus \{\mathbf{0}\} \to [0, \infty)$ such that

$$\lim_{t\to\infty} \frac{D(t\mathbf{x})}{h(t)} = w(\mathbf{x}) \quad \text{for all } \mathbf{x} \neq \mathbf{0}, \tag{3.1.1}$$

and

$$\lim_{t\to\infty} \sup_{\mathbf{u}\in\Theta} \left| \frac{D(t\mathbf{u})}{h(t)} - w(\mathbf{u}) \right| = 0$$

with $0 < \inf_{\mathbf{u}\in\Theta} w(\mathbf{u}) \leq \sup_{\mathbf{u}\in\Theta} w(\mathbf{u}) < \infty$. Moreover, for all $M > 0$, $\inf_{\|\mathbf{x}\|\leq M} D(\mathbf{x}) > 0$.

This is a generalization of the multivariate regular variation condition in Cai et al. (2011), where D is taken as the underlying probability density function. We shall name w the *extreme* depth function. It follows from Assumption D that ξ is unique and w is *homogeneous* of order $-1/\xi$, that is, for all $t > 0$,

$$w(t\mathbf{x}) = t^{-1/\xi} w(\mathbf{x}) \quad \text{for all } \mathbf{x} \neq \mathbf{0}.$$

For many depth functions Assumption D is satisfied directly by construction or under weak conditions on P. Particularly, we discuss the following important examples throughout the paper. It should be emphasized that the applicability of our theory is not limited to only these depth functions.

Example 3.1.1 (Mahalanobis Depth). A very classical example is the Mahalanobis (1936) depth $MD(\mathbf{x}; P) = (1 + d^2_\Sigma(\mathbf{x}, \mu))^{-1}$ with

$$d^2_\Sigma(\mathbf{x}, \mu) = (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$$

where $\Sigma = \Sigma(P)$ is a positive definite $d \times d$ matrix and $\mu = \mu(P) \in \mathbb{R}^d$ is a location parameter. It can be checked that $h(t) = t^{-2}$ with $\xi = \frac{1}{2}$ and $w(\mathbf{x}) = 1/d^2_\Sigma(\mathbf{x}, \mathbf{0})$. As suggested in Liu (1992), we take μ as the mean vector and Σ as the covariance matrix of $\mathbf{X}$ if provided their existence.

Example 3.1.2 (Projection Depth). The projection depth, first considered by Mosteller and Tukey (1977) for a univariate distribution and later generalized to the multivariate case by Donoho and Gasko (1992), is given by $PD_{(\mu,\sigma)}(\mathbf{x}; P) = (1 + O_{(\mu,\sigma)}(\mathbf{x}; P))^{-1}$ with

$$O_{(\mu,\sigma)}(\mathbf{x}; P) = \sup_{\mathbf{u}\in\Theta} \frac{|\mathbf{u}'\mathbf{x} - \mu(F_\mathbf{u})|}{\sigma(F_\mathbf{u})}, \quad \mathbf{x} \in \mathbb{R}^d$$

where $F_\mathbf{u}$ denotes the distribution function of the projection variable $\mathbf{u}^T\mathbf{X}$ and the pair (μ,σ) are given location and scale parameters. Given that

$$\sup_{\mathbf{u}\in\Theta} |\mu(F_\mathbf{u})| < \infty, \quad 0 < \inf_{\mathbf{u}\in\Theta} \sigma(F_\mathbf{u}) \le \sup_{\mathbf{u}\in\Theta} \sigma(F_\mathbf{u}) < \infty, \tag{3.1.2}$$

we can show that $PD(\cdot; P)$ is uniformly continuous on $\mathbb{R}^d$ (Theorem 2.2 in Zuo, 2003) and Assumption D holds with $h(t) = t^{-1}$ ($\xi = 1$) and $w(\mathbf{x}) = 1/O_{(\mathbf{0},\sigma)}(\mathbf{x})$.

Example 3.1.3 (Halfspace Depth). The half-space depth *HD* (Tukey, 1975), one of the most popular choices in the literature, is defined by

$$HD(\mathbf{x}; P) = \inf\{P(H) : \mathbf{x} \in H \in \mathcal{H}\}, \quad \mathbf{x} \in \mathbb{R}^d,$$

where $\mathcal{H}$ is the class of closed half-spaces. It is an important representative from a rich class of the so-called Type-D depth functions defined in Zuo and Serfling (2000a). Assumption D follows by weak multivariate regular variation conditions on the underlying distribution. See Section 2 for more details.

Example 3.1.4 (Spatial Depth). The spatial depth function (Chaudhuri, 1996; Serfling, 2002) is defined by

$$SD(\mathbf{x}; P) = 1 - \left\| E\left\{ \frac{\mathbf{X} - \mathbf{x}}{\|\mathbf{X} - \mathbf{x}\|} \right\} \right\|, \quad \mathbf{x} \in \mathbb{R}^d.$$

Given that $E\|\mathbf{X}\|^2 < \infty$ and the covariance matrix Σ of $\mathbf{X}$ is positive definite, we conjecture that Assumption D holds with $h(t) = t^{-2}$ ($\xi = 1/2$) and $w(\mathbf{x}) = \frac{1}{2\|\mathbf{x}\|^2}\left(\mathrm{tr}\Sigma - \frac{\mathbf{x}'}{\|\mathbf{x}\|}\Sigma\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)$; see Theorem 2 in Girard and Stupfler (2014).

This paper is organized as follows. In Section 3.2, we derive our general extreme estimator of $\mathcal{Q}$ and present its consistency. Section 3.3 presents an asymptotic normality result for constructing naive confidence sets, and Section 3.4 provides a refined asymptotic theory particularly for the half-space depth. A simulation study is carried out in Section 3.5. All the proofs are deferred to Section 3.6. Applications to financial data and some auxiliary results are included in a supplementary document.

3.2 Extreme estimator and its consistency

Consider a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from P and denote the empirical probability measure by $P_n(B) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left[\mathbf{X}_i \in B\right]$ for any Borel set $B \subset \mathbb{R}^d$, where '$\mathbb{1}$' denotes the indicator function. Define the radii $R = \|\mathbf{X}\|$ and $R_i = \|\mathbf{X}_i\|$ for $i = 1, \ldots, n$. We order the R_i's as $R_{1:n} \leq \ldots \leq R_{n:n}$. Denote $F_R(t) = \mathbb{P}(R \leq t)$, $U(t) = F_R^{\leftarrow}\left(1 - \frac{1}{t}\right)$, $t > 0$, where '$\leftarrow$' indicates the left-continuous inverse. For an arbitrary set $B \subset \mathbb{R}^d$ and $t \in \mathbb{R}$, denote $tB = \{t\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \in B\}$.

We shall first derive our extreme estimator of $\mathcal{Q}$ using a so-called *intermediate sequence* $k = k(n) \in \{1, \ldots, n\}$, i.e. we have the following assumption.

Assumption 3.2.1. $k \to \infty$ and $k/n \to 0$, as $n \to \infty$.

Now, using the limit relation (3.1.1) and homogeneity of w, for large

$t = n/k$ we have following approximation

$$\begin{aligned}\mathcal{Q} &= \left\{U\left(n/k\right)\mathbf{z} \in \mathbb{R}^d : D\left(U(n/k)\mathbf{z}; P\right) < \beta\right\} \\ &\approx U\left(n/k\right)\left\{\mathbf{z} \in \mathbb{R}^d : w(\mathbf{z}) < \beta/h(U(n/k))\right\} = U\left(n/k\right)\left(\frac{h(U(n/k))}{\beta}\right)^{\xi} S\end{aligned}$$

where

$$S := \{\mathbf{z} \in \mathbb{R}^d : w(\mathbf{z}) < 1\} = \{\mathbf{z} = r\mathbf{u} : r > (w(\mathbf{u}))^{\xi}, \mathbf{u} \in \Theta\}.$$

The approximation above depends on the depth value β, which is, unfortunately, unknown. To develop a further approximation we need some more regular variation conditions.

Assumption 3.2.2. There exists a so-called extreme value index $\gamma > 0$ and a second order coefficient $\rho < 0$ and a positive or negative function α_R with $\lim_{t\to\infty} \alpha_R(t) = 0$ such that

$$\lim_{t\to\infty} \frac{1}{\alpha_R(t)} \left\{\frac{\mathbb{P}(R > tr)}{\mathbb{P}(R > t)} - r^{-1/\gamma}\right\} = r^{-1/\gamma}\frac{r^{\rho/\gamma} - 1}{\gamma\rho}, \quad r > 0. \tag{3.2.1}$$

This is the standard second-order condition required in the asymptotic theory in *univariate* extreme value theory; see, e.g., Section 2.3 in de Haan and Ferreira (2006). Clearly, it implies the first-order condition that the distribution of R is in the max-domain of attraction of a Fréchet distribution, that is,

$$\lim_{t\to\infty} \frac{\mathbb{P}(R > tr)}{\mathbb{P}(R > t)} = r^{-1/\gamma}, \quad r > 0.$$

Assumption 3.2.3.

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{X} \in tS)}{\mathbb{P}(R > t)} = c \in (0, \infty).$$

This condition requires that the tail probabilities $\mathbb{P}(\mathbf{X} \in tS)$ and $\mathbb{P}(R > t)$ approach zero at the same rate, as $t \to \infty$. It is satisfied by, but not limited to, many multivariate regularly varying distributions; see Assumption R below.

Recall that $\mathcal{Q}$ may be approximated by a proper inflation of S. Assumption 3.2.2 and 3.2.3 then imply that

$$\mathcal{Q} \approx U\left(\frac{c}{p}\right) S \approx U\left(\frac{n}{k}\right)\left(\frac{kc}{np}\right)^{\gamma} S =: \widetilde{\mathcal{Q}}_n, \tag{3.2.2}$$

where the second approximation follows from the regular variation of U; see, e.g., Corollary 1.2.10 in de Haan and Ferreira (2006).

Substituting all components of $\widetilde{\mathcal{Q}}_n$ by their respective estimators yields our extreme estimator of $\mathcal{Q}$. Particularly, we take $\widehat{U}(n/k) = R_{n-k:n}$, the Hill (1975) estimator

$$\widehat{\gamma} = \frac{1}{k}\sum_{i=1}^{k} \log R_{n+1-i:n} - \log R_{n-k:n},$$

and

$$\widehat{c} = \frac{n}{k} P_n(R_{n-k:n}\widehat{S}),$$

with an estimator of S, depending on the choice of depth function, given by

$$\widehat{S} = \{\mathbf{z} = r\mathbf{u} : r > \widehat{\rho}_S(\mathbf{u}), \mathbf{u} \in \Theta\}.$$

for some proper estimator $\widehat{\rho}_S(\cdot)$ of $(w(\cdot))^{\xi}$ on Θ. To conclude, our extreme estimator is given by

$$\widehat{\mathcal{Q}} = \{r\mathbf{u} : r > \widehat{\rho}(\mathbf{u}), \mathbf{u} \in \Theta\}$$

with the function $\widehat{\rho}$ on Θ given by

$$\widehat{\rho}(\mathbf{u}) = R_{n-k:n}\left(\frac{k\widehat{c}}{np}\right)^{\widehat{\gamma}} \widehat{\rho}_S(\mathbf{u}), \quad \mathbf{u} \in \Theta. \tag{3.2.3}$$

Note that $\widehat{\mathcal{Q}}$ can be viewed as an analogue of the Weissman (1978) estimator of an extreme univariate quantile.

Below is a sufficient condition to avoid measurability problems, without which our general result would rely on the outer measure $\mathbb{P}^*$ and the inner measure $\mathbb{P}_*$ when needed.

Assumption M (Measurability)*.* The true $\mathcal{Q} = \mathcal{Q}_n$ and its estimator $\widehat{\mathcal{Q}}$ are both open and their complements are convex and bounded for all $n \in \mathbb{N}$. Moreover, $\widehat{\rho}$ is a stochastic process on Θ with continuous sample paths.

We first provide the following consistency result. It requires a non-trivial construction of $\widehat{\rho}_S$ depending on the choice of depth functions and we shall discuss some interesting examples later on. Below '$\xrightarrow{\mathbb{P}^*}$' denotes the convergence in outer probability; see, e.g., Definition 1.9.1 in van der Vaart and Wellner (1996).

Theorem 3.2.1 (Consistency)**.** *If Assumptions 3.1.1-3.2.3 and D hold and, as* $n \to \infty$*,* $\log(np)/\sqrt{k} \to 0$*, then*

$$\sup_{\mathbf{u}\in\Theta} \left|\widehat{\rho}_S(\mathbf{u}) - (w(\mathbf{u}))^{\xi}\right| \xrightarrow{\mathbb{P}^*} 0 \tag{3.2.4}$$

implies that, as $n \to \infty$*,*

$$\mathbb{P}_* \Big((1+\varepsilon)\mathcal{Q} \subset \widehat{\mathcal{Q}} \subset (1-\varepsilon)\mathcal{Q}\Big) \to 1, \quad \varepsilon > 0. \tag{3.2.5}$$

If further Assumption M holds then, as $n \to \infty$*,*

$$\mathbb{P} \Big((1+\varepsilon)\mathcal{Q} \subset \widehat{\mathcal{Q}} \subset (1-\varepsilon)\mathcal{Q}\Big) \to 1, \quad \varepsilon > 0. \tag{3.2.6}$$

Remark 3.2.1*.* The condition $\log(np)/\sqrt{k} \to 0$ is trivially satisfied if $\lim_{n\to\infty} np \in (0, \infty)$, as $n \to \infty$. For the less extreme case that $np \to \infty$, as $n \to \infty$, the fully nonparametric method may still be employed; see, e.g., Page 788 in Liu et al. (1999).

Remark 3.2.2*.* This consistency result still holds with other proper estimator $\widehat{\gamma}$ of γ such that $\sqrt{k}(\widehat{\gamma} - \gamma) = O_{\mathbb{P}}(1)$; see, e.g., Smith (1987) and Dekkers et al. (1989).

Remark 3.2.3*.* The main consistency results for likelihood depth in Cai et al. (2011) and for half-space depth in He and Einmahl (2016) are consequences of (3.2.5).

In following, without proofs, we provide some natural estimators $\widehat{\rho}_S$ for the Mahalabnobis and projection depth such that satisfying condition (3.2.4) and Assumption M. These two depth functions (and many others) satisfy Assumption D by construction; see Introduction. Similar discussions on half-space depth will follow with more elaborations later on. Below Θ_0 denotes a countable, dense subset of Θ.

Example 3.2.1. For the Mahalanobis depth *MD*, take

$$\widehat{\rho}_S(\mathbf{u}) = (\mathbf{u}^T \widehat{\Sigma}^{-1} \mathbf{u})^{-1/2}, \quad \mathbf{u} \in \Theta,$$

when the sample covariance matrix $\widehat{\Sigma}$ is invertible and otherwise zero (everywhere on Θ). It follows that

$$\widehat{S} = \{\mathbf{z} \in \mathbb{R}^d : \mathbf{z}^T \widehat{\Sigma}^{-1} \mathbf{z} \geq 1\},$$

and therefore $\widehat{\mathcal{Q}}$ is the outer region of a centered ellipsoid. Assumption M and condition (3.2.4) are satisfied for any distribution with a positive definite covariance matrix Σ.

Example 3.2.2. For the projection depth *PD*, take

$$\widehat{\rho}_S(\mathbf{u}) := \left(\sup_{\mathbf{v} \in \Theta_0} \frac{|\mathbf{v}'\mathbf{u}|}{\sigma(F_{n\mathbf{v}})} \right)^{-1}, \quad \mathbf{u} \in \Theta,$$

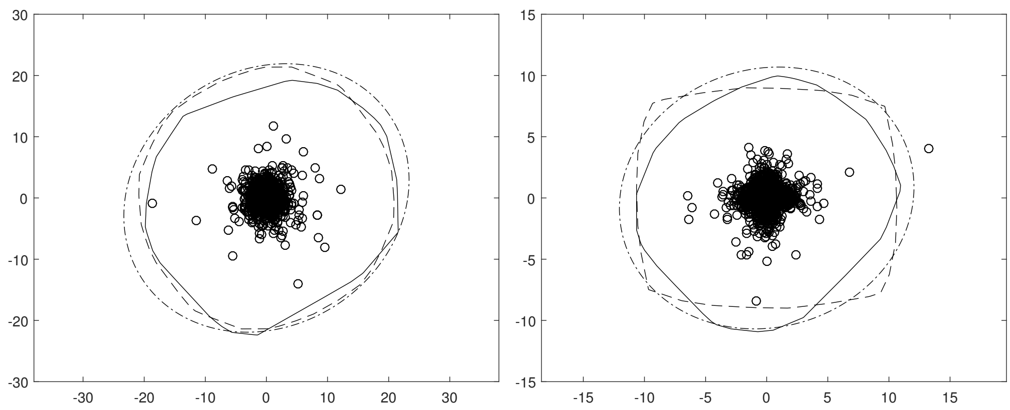
where $F_{n\mathbf{v}}$ is the empirical distribution function of the projected observations $\mathbf{v}'\mathbf{X}_1, \ldots, \mathbf{v}'\mathbf{X}_n$ such that, with probability 1,

$$0 < \inf_{\mathbf{v} \in \Theta_0} \sigma(F_{n\mathbf{v}}) \leq \sup_{\mathbf{v} \in \Theta_0} \sigma(F_{n\mathbf{v}}) < \infty.$$

Assumption M is satisfied with probability 1 if condition (3.1.2) holds and $\sigma(F_{n\mathbf{v}})$ is measurable for all $\mathbf{v} \in \Theta$; see Theorem 2.3 in Zuo (2003). If further provided the continuity of $\sigma(F_\cdot)$ on Θ and

$$\sup_{\mathbf{v} \in \Theta_0} |\sigma(F_{n\mathbf{v}}) - \sigma(F_{\mathbf{v}})| \xrightarrow{\mathbb{P}^*} 0, \quad n \to \infty,$$

then condition (3.2.4) is also satisfied.



(a) Biv. Student t_3 (b) Bivariate Clover

Figure 3.2: Estimated Quantile Regions for $p = 1/1500$ on one sample of size 1500 from the bivariate student t_3 distribution (left) and the bivariate clover distribution (right) with choice of $k = 100$ for the Mahalanobis depth (dash-dotted), projection depth (dashed), and half-space depth (solid).

The regular variation of the half-space depth follows from that of the underlying distribution. In other words, the half-space depth satisfies Assumption D under the following multivariate regular variation condition.

Assumption R. The random vector $\mathbf{X}$ is multivariate regularly varying, that is, there exists a limiting non-zero Radon measure ν such that

$$\frac{\mathbb{P}(\mathbf{X} \in tB)}{\mathbb{P}(R > t)} \to \nu(B) < \infty, \qquad t \to \infty$$

for every Borel set B bounded away from the origin that satisfies $\nu(\partial B) = 0$. In addition, let $\nu(B) > 0$ if $B \supset H$ for some $H \in \mathcal{H}$.

This is a multivariate generalization of the so-called 'peaks-over-threshold' method in univariate extreme value theory; see, e.g., Section 5.4 in Resnick (2007). Assumption R is satisfied by many multivariate heavy-tailed distributions, including the heavy-tailed elliptical class (Hashorva, 2006). It is well known that ν is homogeneous, that is, for all $t > 0$

$$\nu(tB) = t^{-1/\gamma}\nu(B),$$

see, e.g., de Haan and Resnick (1979). Clearly, ν defines a probability measure on $\mathbb{C} = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| > 1\}$. For the half-space depth, Assumption R implies Assumption D with $h(t) = \mathbb{P}(R > t)$ (so $\xi = \gamma$), and the *extreme* half-space depth function given by

$$w(\mathbf{z}) := HD(\mathbf{z}; \nu) := \inf \{\nu(H) : \mathbf{z} \in H \in \mathcal{H}\}, \quad \mathbf{z} \in \mathbb{R}^d.$$

Observe that Assumption 3.2.3 is also satisfied with $c = \nu(S)$.

Example 3.2.3. For the half-space depth HD, take

$$\widehat{\rho}_S(\mathbf{u}) = \left(\widehat{HD}(\mathbf{u}; \widehat{\nu}^*)\right)^{\widehat{\gamma}} = \left(\inf_{\mathbf{v}\in\Theta_0} \widehat{\nu}^* \left(H_{\mathbf{u}^T\mathbf{v},\mathbf{v}}\right)\right)^{\widehat{\gamma}}, \quad \mathbf{u} \in \Theta,$$

with $\widehat{\nu}^*$ the 'normalized' estimator of ν on half-spaces given by

$$\widehat{\nu}^*(H_{r,\mathbf{u}}) = (r \vee k^{-\widehat{\gamma}})^{-1/\widehat{\gamma}} \cdot \frac{n}{k} P_n \left(R_{n-k:n} H_{1,\mathbf{u}}\right),$$

and

$$H_{r,\mathbf{u}} := \{\mathbf{x} \in \mathbb{R}^d : \mathbf{u}^T\mathbf{x} \geq r\}, \quad (r, \mathbf{u}) \in \mathbb{R} \times \Theta.$$

Assumption M then immediately holds; see Theorem 2.11 in Zuo and Serfling (2000a) and Proposition 3 in He and Einmahl (2016). Condition (3.2.4) is satisfied under Assumption R, by Lemma 7 in He and Einmahl (2016) and the consistency of $\widehat{\gamma}$ from, e.g., Theorem 3.2.2 in de Haan and Ferreira (2006).

Figure 3.2 compares the estimated quantile regions for $p = 1/1500$ with choice of $k = 100$ for the Mahalanobis depth (dash-dotted), projection depth (dashed) and half-space depth (solid) based on one sample of size 1500 respectively from the bivariate student t_3 distribution and the bivariate clover distribution (see simulation section 3.5 for definition). For the bivariate student distribution, the true extreme regions coincide for all three depth functions, and we can see the estimated quantile regions are close to each other. For the bivariate clover distribution, our estimates suggest that the shape of the population quantile regions seem to be quite different for the considered depth functions.

3.3 Asymptotic Normality

To establish the asymptotic normality of our extreme estimator of $\mathcal{Q}$, we need the following extra second-order conditions.

Assumption 3.3.1. Either $h(t) = c_h t^{-1/\xi}$ for some constant $c_h > 0$, or there exist $\rho' < 0$ and a positive or negative function α_h with $\lim_{t\to\infty} \alpha_h(t) = 0$ such that

$$\lim_{t\to\infty} \frac{1}{\alpha_h(t)} \left(\frac{h(tr)}{h(t)} - r^{-1/\xi} \right) = r^{-1/\xi} \frac{r^{\rho'/\xi} - 1}{\xi\rho'}, \quad r > 0. \tag{3.3.1}$$

Assumption 3.3.2. For some positive function α regularly varying at infinity with negative index, as $t \to \infty$,

$$\sup_{\mathbf{u}\in\Theta} \left| \frac{D(t\mathbf{u})}{h(t)} - w(\mathbf{u}) \right| = O(\alpha(t)), \quad \frac{\mathbb{P}(\mathbf{X} \in tS)}{\mathbb{P}(R \geq t)} - c = O(\alpha(t)).$$

We take α such that $|\alpha_h(t)| \leq \alpha(t)$.

Condition (3.3.1) is a standard second-order condition, analogous to (3.2.1), in the theory of regularly varying functions; see, e.g., Appendix B.3 in de Haan and Ferreira (2006). For the half-space depth, it is just a restatement of (3.2.1); see the end of Section 2 above. Assumption 3.3.2 quantifies the rates of convergence in Assumption D and Assumption 3.2.3.

For presentation convenience, we introduce the notion of (directed) logarithmic distance between our true and estimated quantile regions. Precisely, for arbitrary two sets $B, \widetilde{B} \subset \mathbb{R}^d$, we define their *logarithmic distance* by

$$\Delta_2(B, \widetilde{B}) = \Delta_2(\widetilde{B}, B) = \inf \left\{ \delta \in \mathbb{R}_+ : e^{\delta}\widetilde{B} \subset B \subset e^{-\delta}\widetilde{B} \right\} \in [0, \infty],$$

and, similarly, their *directed logarithmic distance* by

$$\Delta_1(B, \widetilde{B}) = \inf \left\{ \delta \in \mathbb{R} : B \subset e^{-\delta}\widetilde{B} \right\} \in [-\infty, \infty]$$

Since $a\widehat{\mathcal{Q}} \subset a'\widehat{\mathcal{Q}}$ if $0 \leq a' \leq a \leq \infty$, it is easy to verify for all $\delta \in (0, \infty)$ that

$$\Delta_2(\mathcal{Q}, \widehat{\mathcal{Q}}) \leq \delta \Leftrightarrow e^{\delta}\widehat{\mathcal{Q}} \subset \mathcal{Q} \subset e^{-\delta}\widehat{\mathcal{Q}},$$

$$\Delta_1(\mathcal{Q}, \widehat{\mathcal{Q}}) \le \delta \Leftrightarrow \mathcal{Q} \subset e^{-\delta}\widehat{\mathcal{Q}}, \quad \Delta_1(\widehat{\mathcal{Q}}, \mathcal{Q}) \le \delta \Leftrightarrow \widehat{\mathcal{Q}} \subset e^{-\delta}\mathcal{Q}.$$

Denote $A_R(t) = \alpha_R(U(t))$ and $A(t) = \alpha(U(t))$.

Theorem 3.3.1 (Asymptotic Normality). *Given Assumptions 3.1.1-3.3.2 and D and, as* $n \to \infty$, $np = o(k)$, $\sqrt{k}A_R(n/k) \to 0$, $\frac{\sqrt{k}}{\log(k/(np))}A(n/k) \to 0$, *then*

$$\max\left\{\frac{\sqrt{k}}{\log(k/(np))}, 1\right\} \sup_{\mathbf{u}\in\Theta} \left|\widehat{\rho}_S(\mathbf{u}) - (w(\mathbf{u}))^{\xi}\right| \xrightarrow{\mathbb{P}^*} 0 \tag{3.3.2}$$

implies that

$$\frac{\sqrt{k}}{\log(k/(np))}\Delta_1(\mathcal{Q}, \widehat{\mathcal{Q}}) - \Gamma_n \xrightarrow{\mathbb{P}^*} 0, \frac{\sqrt{k}}{\log(k/(np))}\Delta_1(\widehat{\mathcal{Q}}, \mathcal{Q}) + \Gamma_n \xrightarrow{\mathbb{P}^*} 0, \tag{3.3.3}$$

$$\frac{\sqrt{k}}{\log(k/(np))}\Delta_2(\mathcal{Q}, \widehat{\mathcal{Q}}) - |\Gamma_n| \xrightarrow{\mathbb{P}^*} 0 \tag{3.3.4}$$

where $\Gamma_n := \sqrt{k}(\widehat{\gamma} - \gamma) \xrightarrow{d} N(0, \gamma^2)$. *The convergences in* (3.3.3) *and* (3.3.4) *hold with the probability measure* $\mathbb{P}$ *replacing* $\mathbb{P}^*$ *if Assumption M is satisfied.*

Remark 3.3.1. The condition $\sqrt{k}A_R(n/k) \to 0$, $n \to \infty$, is imposed to eliminate the asymptotic bias of the Hill estimator; see Theorem 3.2.5 in de Haan and Ferreira (2006). In the more general case $\sqrt{k}A_R(n/k) \to \lambda$ with λ finite, Theorem 3.3.1 holds with $\Gamma_n \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right)$. With other choices of $\widehat{\gamma}$ such that $\Gamma_n = O_{\mathbb{P}}(1)$, see Remark 3.2.2, (3.3.3) an (3.3.4) still hold under the same conditions without requiring $\sqrt{k}A_R(n/k) \to 0$, as $n \to \infty$.

Remark 3.3.2. The condition $\log(np)/\sqrt{k} \to 0$, as $n \to \infty$, is not necessary for Theorem 3.3.1. This suggests that Theorem 3.3.1 can still hold even if $\widehat{\mathcal{Q}}$ is inconsistent in the sense of Theorem 3.2.1.

Base on following Corollary, we can easily construct (simultaneous) asymptotic confidence sets of $\mathcal{Q}$ or/and $\mathcal{Q}^c$. They will be called *naive* since only the estimation error of the extreme value index γ plays a role in the limits in Theorem 3.2.1.

Corollary 3.3.1 (Naive Asymptotic Confidence Sets)**.** *For all* $\tau \in [0,1]$ *define the events*

$$\begin{aligned} C_\tau^+ &= \left\{ \mathcal{Q} \subset \exp\left(-\frac{\log(k/(np))}{\sqrt{k}} \widehat{\gamma} \Phi^{-1}(\tau) \right) \widehat{\mathcal{Q}} \right\}, \\ C_\tau^- &= \left\{ \exp\left(\frac{\log(k/(np))}{\sqrt{k}} \widehat{\gamma} \Phi^{-1}(\tau) \right) \widehat{\mathcal{Q}} \subset \mathcal{Q} \right\}, \end{aligned}$$

where Φ^{-1} *denotes the standard normal quantile function. Under the conditions of Theorem 3.3.1 and Assumption M, for all* $\tau \in (0,1)$*, as* $n \to \infty$*,*

$$\mathbb{P}(C_\tau^+) \to \tau,\ \mathbb{P}(C_\tau^-) \to \tau,\ \mathbb{P}(C^+_{(1+\tau)/2} \cap C^-_{(1+\tau)/2}) \to \tau. \tag{3.3.5}$$

Remark 3.3.3. Under the conditions of Theorem 3.3.1 without Assumption M, (3.3.5) holds with either $\mathbb{P}^*$ or $\mathbb{P}_*$ replacing $\mathbb{P}$ for all $\tau \in (0,1)$.

Condition (3.3.2) can be easily verified for the Mahalanobis depth with $\widehat{\rho}_S$ given in Corollary 3.2.1 when the underlying distribution has a positive definite covariance matrix. It is also satisfied by the extreme estimator for the projection depth, under the conditions of Corollary 3.2.2 in conjunction with that, as $n \to \infty$,

$$\sup_{\mathbf{v} \in \Theta_0} |\sigma(F_{n\mathbf{v}}) - \sigma(F_{\mathbf{v}})| = O_{\mathbb{P}}(1/\sqrt{n}). \tag{3.3.6}$$

For example, for the pair (μ, σ)=(mean, standard deviation), (3.3.6) holds for any underlying distribution with a positive definite covariance matrix. For the pair (μ, σ)=(Med, MAD) where 'Med' denotes univariate median and 'MAD' denotes the univariate median absolute deviation, (3.3.6) is satisfied, for instance, when P is a smooth elliptically symmetric distribution; see Remark 2.4 and Remark 3.3 in Zuo (2003).

3.4 Refined asymptotics for the half-space depth

Since the asymptotic limits in Theorem 3.3.1 do not involve the *shape* estimation uncertainty of $\mathcal{Q}$, the *naive* asymptotic confidence sets might lead to unsatisfying finite-sample coverage probability if the shape of $\mathcal{Q}$ is not estimated effectively. The possibility of this problem is clearly demonstrated for the half-space depth in our simulation study in the next section. This motivates a *refined* asymptotic theory to take the shape estimation error into account, at least, for the half-space depth. Henceforth we shall always suppose Assumption R holds, i.e. the underlying distribution is multivariate regularly varying. Recall that Assumption D is satisfied if further provided Assumption 3.2.2.

3.4.1 Adjusted extreme estimator and asymptotically conservative confidence sets

We need another intermediate sequence $k_1 = k_1(n) \in \{1, \ldots, n\}$.

Assumption 3.4.1. $k_1/n \to 0$ and $k_1 \to \infty$, as $n \to \infty$.

We introduce another estimator of $HD(\mathbf{u}; \nu)$, using k_1 instead of k (see Corollary 3.2.3 for comparison), given by

$$\widehat{HD}(\mathbf{u}; \widehat{\nu}_1^*) = \inf_{\mathbf{v} \in \Theta_0} \widehat{\nu}_1^*(H_{\mathbf{u}^T \mathbf{v}, \mathbf{v}}), \quad \mathbf{u} \in \Theta,$$

with another 'normalized' estimator of the exponent measure

$$\widehat{\nu}_1^*(H_{r,\mathbf{u}}) := \left(r \vee k_1^{-\widehat{\gamma}}\right)^{-1/\widehat{\gamma}} \cdot \frac{n}{k_1} P_n\left(R_{n-k_1:n} H_{1,\mathbf{u}}\right), \quad (r, \mathbf{u}) \in \mathbb{R} \times \Theta.$$

We also propose an *affine-invariance* estimator of $c = \nu(S)$ given by

$$\widehat{c}_1 = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left[\widehat{HD}(\mathbf{X}_i; P_n) \leq \frac{k}{n}\right] = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left[\inf_{\mathbf{v} \in \Theta_0} P_n(H_{\mathbf{v}^T \mathbf{X}_i, \mathbf{v}}) \leq \frac{k}{n}\right].$$

Our adjusted extreme estimator of $\mathcal{Q}$ is then given by

$$\widehat{\mathcal{Q}}_1 = \{r\mathbf{u} : r > \widehat{\rho}_1(\mathbf{u}), \mathbf{u} \in \Theta\},$$

with the adjusted radius function, cf. (3.2.3),

$$\widehat{\rho}_1(\mathbf{u}) := R_{n-k_1:n} \left(\frac{k_1 \widehat{c}_1}{np}\right)^{\widehat{\gamma}} \left(\widehat{HD}(\mathbf{u}; \widehat{\nu}_1^*)\right)^{\widehat{\gamma}}, \quad \mathbf{u} \in \Theta.$$

We quantifies the rate of convergence in Assumption R with following condition, which in conjunction with Assumption 3.2.2 imply Assumption 3.3.2.

Assumption 3.4.2. For some positive function α regularly varying at infinity with negative index, as $t \to \infty$,

$$\frac{\mathbb{P}(\mathbf{X} \in tH)}{\mathbb{P}(R > t)} - \nu(H) = O(\alpha(t)),$$

uniformly for $H \in \{H_{1,\mathbf{u}} : \mathbf{u} \in \Theta\} \cup \{S\}$. We take α such that $|\alpha_R(t)| \leq \alpha(t)$.

Recall that $A(\cdot) = \alpha(U(\cdot))$ and $A_R(\cdot) = \alpha_R(U(\cdot))$. The following theorem allows the construction of asymptotically *conservative* confidence sets.

Theorem 3.4.1 (Asymptotically Conservative Confidence Sets)**.** *Let*

$$\lim_{n\to\infty} \psi_n := \lim_{n\to\infty} \frac{\sqrt{k}}{\sqrt{k_1}\log(k_1/(np))} =: \psi < \infty.$$

For all $\tau \in [0, 1]$, define the events

$$\bar{C}_\tau^+ = \left\{ \mathcal{Q} \subset \exp\left(-\frac{\gamma q_1(\tau)}{\sqrt{k_1}\psi_n}\right) \widehat{\mathcal{Q}}_1 \right\},\ \bar{C}_\tau^- = \left\{ \exp\left(\frac{\gamma q_1(\tau)}{\sqrt{k_1}\psi_n}\right) \widehat{\mathcal{Q}}_1 \subset \mathcal{Q} \right\}$$
$$\bar{C}_\tau = \left\{ \exp\left(\frac{\gamma q_2(\tau)}{\sqrt{k_1}\psi_n}\right) \widehat{\mathcal{Q}}_1 \subset \mathcal{Q} \subset \exp\left(-\frac{\gamma q_2(\tau)}{\sqrt{k_1}\psi_n}\right) \widehat{\mathcal{Q}}_1 \right\},$$

where q_1 and q_2 are, respectively, the quantile functions of

$$Z_1 = \sup_{\mathbf{u}\in\Theta} \mathbb{Z}(\mathbf{u}), \quad Z_2 = \sup_{\mathbf{u}\in\Theta} |\mathbb{Z}(\mathbf{u})|$$

and $\mathbb{Z}$ is a mean-zero Gaussian process on Θ with covariance structure

$$Cov(\mathbb{Z}(\mathbf{u}), \mathbb{Z}(\mathbf{v})) = 1 + \psi^2 \cdot \frac{\nu(H_{1,\mathbf{u}} \cap H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{u}})\nu(H_{1,\mathbf{v}})} \quad \mathbf{u}, \mathbf{v} \in \Theta. \tag{3.4.1}$$

Given Assumptions 3.1.1-3.2.2, R, 3.4.1, 3.4.2 and, as $n \to \infty$, $np = o(k_1)$, $\sqrt{k_1}A_R(n/k_1) \to 0$, $\sqrt{k}A_R(n/k) \to 0$, $\frac{\sqrt{k}}{\log(k_1/(np))}A(n/k_1) \to 0$, $\frac{\sqrt{k}}{\log(k_1/(np))}A(n/k) \to 0$, it holds for all $\tau \in [0, 1]$ that

$$\liminf_{n\to\infty} \mathbb{P}\left(\bar{C}_\tau^+\right) \geq \tau,\ \liminf_{n\to\infty} \mathbb{P}\left(\bar{C}_\tau^-\right) \geq \tau,\ \liminf_{n\to\infty} \mathbb{P}\left(\bar{C}_\tau\right) \geq \tau.$$

To construct asymptotically *conservative* confidence sets of $\mathcal{Q}$ or $\mathcal{Q}^c$ or both simultaneously based on Theorem 3.4.1, one need consistent estimators of γ and the corresponding quantile of Z_1 or Z_2. In a finite sample, this may be achieved with $\widehat{\gamma}$, and a sample version of $\mathbb{Z}$, denoted as $\mathbb{Z}_n$, with the estimated covariance structure

$$\mathbb{E}\left(\mathbb{Z}_n(\mathbf{u}) \cdot \mathbb{Z}_n(\mathbf{v})|\mathbf{X}_1, \ldots, \mathbf{X}_n\right) = 1 + \psi_n^2 \cdot \frac{\widehat{\nu}(H_{1,\mathbf{u}} \cap H_{1,\mathbf{v}})}{\widehat{\nu}(H_{1,\mathbf{u}})\widehat{\nu}(H_{1,\mathbf{v}})} \quad \mathbf{u}, \mathbf{v} \in \Theta,$$

where $\widehat{\nu}(B) = \frac{n}{k} P_n(R_{n-k:n} B)$, for any Borel set B.

In the following subsection we shall show that these asymptotically conservative confidence sets are also asymptotically *correct* with few additional conditions, which can easily satisfied by many heavy-tailed (elliptical) distributions such as those in our simulation study below.

3.4.2 Refined Asymptotic Confidence Sets

In this subsection, we provide a refined asymptotic theory for constructing asymptotically correct confidence sets mainly of theoretical interest. Generally speaking, this involves a delicate estimation of the so-called *minimal directions* (in terminology of Massé, 2004) in the calculation of the extreme half-space depth on the unit sphere. Precisely, we need to estimate a map, say, ϕ from the unit sphere Θ to its power set $\mathcal{P}(\Theta)$ such that

$$\phi(\mathbf{u}) := \left\{\mathbf{v} \in \Theta : \nu\left(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}\right) = HD(\mathbf{u}; \nu)\right\}, \quad \mathbf{u} \in \Theta.$$

We can show that $\phi(\mathbf{u})$ is nonempty, closed and $\phi(\mathbf{u}) \subset \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma\}$ with $\delta_0 := \inf_{\mathbf{u} \in \Theta} \nu(H_{1,\mathbf{u}}) > 0$ for all $\mathbf{u} \in \Theta$; see Lemma 3.8.6 below.

Assumption 3.4.3. For all $\delta > 0$,

$$\inf_{\mathbf{u} \in \Theta} \inf_{\mathbf{w} \in \Theta_0 \setminus \phi(\mathbf{u}, \delta)} \left\{\nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u}; \nu)\right\} > 0,$$

where $\phi(\mathbf{u}, \delta) = \{\mathbf{w} \in \Theta : \inf_{\mathbf{v} \in \phi(\mathbf{u})} \|\mathbf{w} - \mathbf{v}\| \leq \delta\}$.

This condition requires that the extreme half-space depth value on the entire unit sphere cannot be achieved in the directions outside neighborhoods of ϕ. It is a tail analogue of a crucial condition in Theorem 4.1 in Arcones et al. (2006) for developing the asympotics of the empirical half-space depth process.

Theorem 3.4.2 (Refined Asymptotics)**.** *Under the conditions of Theorem 3.4.1 in conjunction with Assumption 3.4.3, as $n \to \infty$,*

$$\frac{\sqrt{k}}{\log(k_1/(np))} \begin{pmatrix} \Delta_1(\mathcal{Q}, \widehat{\mathcal{Q}}_1) \\ \Delta_1(\widehat{\mathcal{Q}}_1, \mathcal{Q}) \\ \Delta_2(\mathcal{Q}, \widehat{\mathcal{Q}}_1) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \sup_{\mathbf{u}\in\Theta} \inf_{\mathbf{v}\in\phi(\mathbf{u})} \mathbb{Z}(\mathbf{v}) \\ \sup_{\mathbf{u}\in\Theta} \sup_{\mathbf{v}\in\phi(\mathbf{u})} -\mathbb{Z}(\mathbf{v}) \\ \sup_{\mathbf{u}\in\Theta} \left|\inf_{\mathbf{v}\in\phi(\mathbf{u})} \mathbb{Z}(\mathbf{v})\right| \end{pmatrix}$$

with $\mathbb{Z}$ as in Theorem 3.4.1.

Generally speaking, the *refined* asymptotic confidence sets of $\widehat{\mathcal{Q}}$ or $\widehat{\mathcal{Q}}^c$, or both simultaneously, may be constructed based Theorem 3.4.2 with $\widehat{\gamma}$, $\mathbb{Z}_n$, and a possible approximation of ϕ based on $\widehat{\nu}_1^*$ given by

$$\widehat{\phi}_n(\mathbf{u}) = \left\{\mathbf{w} \in \Theta : \widehat{\nu}_1^* \left(H_{\mathbf{w}^T\mathbf{u},\mathbf{w}}\right) = \widehat{HD}(\mathbf{u}; \widehat{\nu}_1^*)\right\}, \quad \mathbf{u} \in \Theta.$$

From a practical point of view, such an approximation of ϕ may be unfortunately too noisy to be adopted. To avoid this, one should notice that Theorem 3.4.2 implies that, under the conditions of the following corollary, the asymptotically conservative confidence sets can (and will) be taken as our *refined* asymptotic confidence sets without estimating ϕ. Note that Assumption 3.4.3 is satisfied if $\phi(\mathbf{u})$ is a singleton for all $\mathbf{u} \in \Theta$; see Lemma 3.8.7 in the supplementary document.

Corollary 3.4.1. *Let $\phi(\mathbf{u})$ be a singleton for all $\mathbf{u} \in \Theta$, and $\cup_{\mathbf{u}\in\Theta}\phi(\mathbf{u}) = \Theta$. Under the conditions of Theorem 3.4.1, as $n \to \infty$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\bar{C}_\tau^+\right) = \tau, \; \lim_{n\to\infty} \mathbb{P}\left(\bar{C}_\tau^-\right) = \tau, \; \lim_{n\to\infty} \mathbb{P}\left(\bar{C}_\tau\right) = \tau.$$

3.5 Simulation study

In this section, a simulation study is carried out to examine the performance of our asymptotic approximations in finite samples with size $n = 1500$ for the half-space depth and the projection depth. Throughout this section a probability value $p = 1/n$ is taken for both the quantile regions and central regions. For the projection depth, we choose the pair (μ, σ)=(Med, MAD), where 'Med' denotes univariate median and 'MAD' denotes median absolute deviation. We consider five multivariate distributions as follows. Unless specified otherwise, in this section we shall always take $k = k_1 = 100$ (with $\psi_n \approx 0.312$).

- The bivariate Cauchy distribution ($\gamma = 1$) with density

$$f(x, y) = \frac{1}{2\pi} \frac{1}{(1 + (x^2 + y^2))^{3/2}}, \qquad (x, y) \in \mathbb{R}^2$$

- The Student t_3 distribution ($\gamma = 1/3$) with density

$$f(x, y) = \frac{1}{2\pi} \frac{1}{(1 + (x^2 + y^2)/3)^{5/2}}, \qquad (x, y) \in \mathbb{R}^2$$

- An affine Cauchy distribution ($\gamma = 1$)

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}, \quad \mathbf{A} = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 1 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

  where $\mathbf{Y}$ has the bivariate Cauchy distribution defined above.

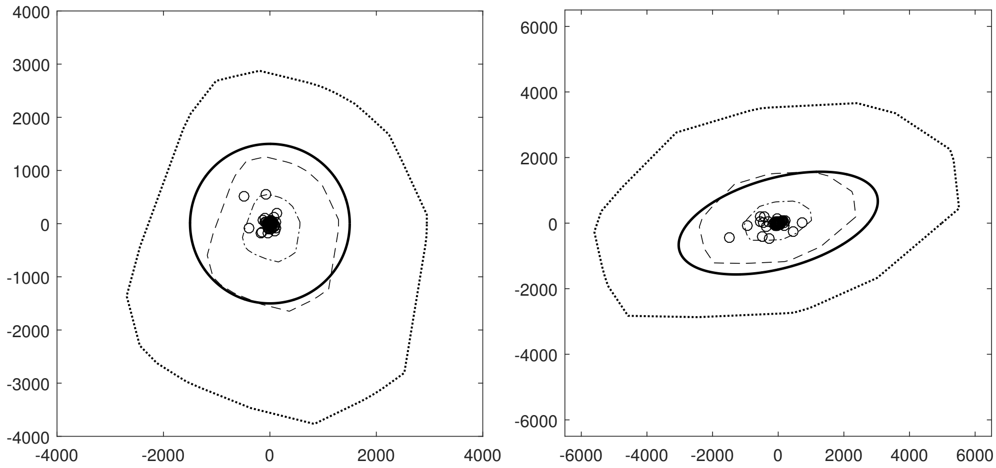
- The trivariate Cauchy distribution ($\gamma = 1$) with density

$$f(x) = \frac{1}{\pi^2 (1 + x^2 + y^2 + z^2)^2}, \qquad (x, y, z) \in \mathbb{R}^3.$$

- The bivariate "clover" distribution ($\gamma = 1/3$) with density

$$f(x, y) = \frac{3\left(9(x^2 + y^2)^2 - 32x^2y^2\right)}{10\pi\left(1 + (x^2 + y^2)^3\right)^{3/2}}, \quad x, y \in \mathbb{R}.$$

This is a distribution with clover-shaped (hence non-elliptical and non-convex) density contours. Recall, however, that both half-space depth and projection-depth based central regions are convex.

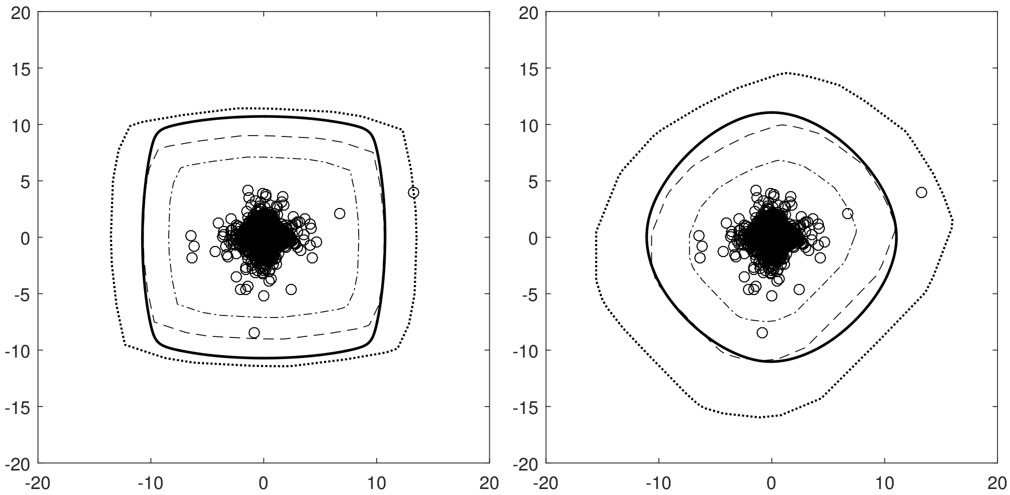


(a) Biv. Cauchy (b) Aff. Biv. Cauchy

Figure 3.3: True (solid) and estimated (dashed) half-space depth based quantile regions and simultaneous 75% refined asymptotic confidence sets of the quantile regions (open outer region of the dash-dotted line) and the central regions (closed inner region of the dotted line) at level $p = 1/n$ based on one sample of size $n = 1500$ for $k = k_1 = 100$.

The first four (elliptical) distributions satisfy all the distributional assumptions in Corollaries 3.3.1 and 3.4.1 for both the projection and half-space depth functions: the (simultaneous) asymptotically conservative confidence sets from Section 3.4.1 therefore will be taken as our (simultaneous) *refined* asymptotic confidence sets. Although it is difficult to verify all these assumptions for the clover distribution, we shall see nevertheless satisfactory finite-sample coverage probabilities of our refined asymptotic ‘confidence’ sets.

Recall Figure 3.1 in Introduction showing the true and estimated quantile regions, and simultaneous 75% refined asymptotic confidence sets of the



(a) Projection Depth (b) Half-space Depth

Figure 3.4: True (solid) and estimated (dashed) quantile regions for the projection depth (left) and half-space depth (right), and simultaneous naive (left) and refined (right) 90% asymptotic 'confidence' sets of the quantile regions (open outer region of the dash-dotted line) and central regions (closed inner region of the dotted line) at level $p = 1/n$ based on one common sample from the bivariate clover distribution of size $n = 1500$ for $k = k_1 = 100$.

half-space depth based quantile region $\mathcal{Q}$ and central region $\mathcal{Q}^c$ based on one sample from the bivariate student t_3 distribution. Figure 3.3 contains two similar plots each based on one sample respectively for the bivariate Cauchy distribution and our affine Cauchy distribution. The left plot in Figure 3.4 presents our extreme estimates and naive asymptotic confidence sets for the projection depth based $\mathcal{Q}$ and $\mathcal{Q}^c$ based on one sample from the clover distribution. We also provide the right plot using the half-space depth for comparison, in which we present the refined asymptotic confidence sets instead of the naive ones. Here we can depict only approximate true quantile regions because of computational complexity. Our extreme estimates are all close to the true quantile regions, which are all covered by their respective asymptotic confidence sets.

Table 3.1: The finite-sample coverage probability (in %) of the (simultaneous) 90% naive asymptotic confidence sets of the projection depth based quantile and central regions at level $p = 1/n$ based on 1000 samples of size $n = 1500$ for $k = 100$.

	Both	Central	Quantile
$\gamma = 1$			
Biv. Cauchy	88.3	89.1	86.6
Affine Cauchy	87.2	86.1	90.3
Triv. Cauchy	84.8	83.6	88.3
$\gamma = 1/3$			
Biv. Student t_3	77.6	87.1	75.4
Biv. Clover	71.9	76.9	80.6

Table 3.1 presents the finite-sample coverage probability (in percentage) of the 90% naive asymptotic confidence sets of the projection depth based quantile and central regions in our 1000 simulated samples. The true confidence level is better matched for the distributions with heavier tails, for which the shape estimation error is less weighted in large samples.

Table 3.2 shows the finite-sample coverage probability (in percentage) of both the naive and refined (simultaneous) asymptotic confidence sets of the half-space depth based quantile and central regions at the confidence levels 90% and 75% in 1000 simulated samples. Apparently, the naive asymptotic confidence sets can hardly capture the actual statistical uncertainty in these samples and the refined ones match the true confidence levels much better.

3.6 Proofs

Define the function $V(t) := h^{\leftarrow}(1/t)$, $t > 0$, and recall that $\mathbb{C} := \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| > 1\}$ and $\delta_0 := \inf_{\mathbf{u}\in\Theta} \nu(H_{1,\mathbf{u}}) > 0$. We shall first prove Theorem 3.3.1,

Table 3.2: The finite-sample coverage probability (in %) of the (simultaneous) 90% and 75% naive and refined asymptotic confidence sets of the half-space depth based quantile and central regions at level $p = 1/n$ based on 1000 samples of size $n = 1500$ for $k = k_1 = 100$.

	Naive						Refined					
Conf. Set	Both		Central		Quantile		Both		Central		Quantile	
Conf. Level (%)	90	75	90	75	90	75	90	75	90	75	90	75
Biv. Cauchy	50.4	25.7	64.9	47.9	66.5	50.2	90.7	76.1	87.0	73.5	94.0	83.7
Biv. Student t_3	30.0	10.4	39.8	24.5	74.4	57.3	90.8	78.3	90.3	78.3	92.9	75.9
Affine Cauchy	40.0	18.7	62.0	45.5	61.5	44.8	90.9	77.6	85.9	70.1	95.3	85.8
Clover	31.1	8.4	52.9	37.4	57.0	40.4	90.7	77.5	87.8	74.8	93.9	80.3
Triv. Cauchy	47.7	22.0	60.1	44.0	68.5	50.7	87.2	73.3	82.0	66.4	95.0	86.8

and then Theorem 3.2.1, and lastly Theorems 3.4.1 and 3.4.2.

Lemma 3.6.1. *Under Assumptions D, 3.3.1 and 3.3.2, for all $\varepsilon > 0$, as $t \to \infty$,*

$$\sup_{\|\mathbf{x}\| \geq \varepsilon} \left| \frac{D(t\mathbf{x})}{h(t)} - w(\mathbf{x}) \right| = O(\alpha(t)).$$

Proof. For sufficiently large t,

$$\begin{aligned}
\sup_{\|\mathbf{x}\| \geq \varepsilon} \left| \frac{D(t\mathbf{x})}{h(t)} - w(\mathbf{x}) \right| &= \sup_{r \geq \varepsilon, \mathbf{u} \in \Theta} \left| \frac{D(tr\mathbf{u})}{h(t)} - r^{-1/\xi} w(\mathbf{u}) \right| \\
&\leq \sup_{r \geq \varepsilon} r^{-1/\xi} \sup_{r \geq \varepsilon, \mathbf{u} \in \Theta} \left| \frac{D(tr\mathbf{u})}{h(tr)} - w(\mathbf{u}) \right| + \sup_{r \geq \varepsilon, \mathbf{u} \in \Theta} \frac{D(tr\mathbf{u})}{h(tr)} \sup_{r \geq \varepsilon} \left| \frac{h(tr)}{h(t)} - r^{-1/\xi} \right| \\
&\leq \varepsilon^{-1/\xi} \sup_{r \geq \varepsilon, \mathbf{u} \in \Theta} \left| \frac{D(tr\mathbf{u})}{h(tr)} - w(\mathbf{u}) \right| + 2 \sup_{\mathbf{u} \in \Theta} w(\mathbf{u}) \sup_{r \geq \varepsilon} \left| \frac{h(tr)}{h(t)} - r^{-1/\xi} \right| =: T_1 + T_2.
\end{aligned}$$

Applying Proposition B.1.10 in de Haan and Ferreira (2006) with Assumption 3.3.2 yields that

$$\sup_{r \geq \varepsilon, \mathbf{u} \in \Theta} \left| \frac{D(tr\mathbf{u})}{h(tr)} - w(\mathbf{u}) \right| = O(\sup_{r \geq \varepsilon} \alpha(tr)) = O(\alpha(t)),$$

and therefore $T_1 = O(\alpha(t))$. It remains to check that $T_2 = O(\alpha(t))$, as well. This is trivial if $h(t) = c_h t^{-1/\xi}$ for some $c_h > 0$ since then $T_2 \equiv 0$. Otherwise, by Theorem 2.3.9 in de Haan and Ferreira (2006), there exists a function $\widetilde{\alpha}_h(t) \sim \alpha_h(t)$ such that when t is sufficiently large for all $r \geq \varepsilon$ we have

$$\left| \frac{1}{\widetilde{\alpha}_h(t)} \left\{ \frac{h(tr)}{h(t)} - r^{-1/\xi} \right\} - r^{-1/\xi} \frac{r^{\rho'/\xi} - 1}{\rho' \xi} \right| \leq r^{-1/\xi + \rho'/\xi} \max\{r^{1/(2\xi)}, r^{-1/(2\xi)}\}.$$

Rearranging this inequality implies that

$$\sup_{r \geq \varepsilon} \left| \frac{h(tr)}{h(t)} - r^{-1/\xi} \right| = O(\widetilde{\alpha}_h(t)) = O(\alpha_h(t)) = O(\alpha(t)). \tag{3.6.1}$$

Hence, $T_2 = O(\alpha(t))$ since $\sup_{\mathbf{u} \in \Theta} w(\mathbf{u}) \in (0, \infty)$ by Assumption D. $\square$

Lemma 3.6.2. *Under Assumption D, for all $\varepsilon > 0$, there exists a $\delta > 0$ and $t_0 > 0$ such that for $t > t_0$,*

$$\left\{ \mathbf{z} \in \mathbb{R}^d : \frac{D(t\mathbf{z})}{h(t)} \leq \varepsilon \right\} \subset \left\{ \mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| > \delta \right\} = \delta \mathbb{C}.$$

The proof is analogous as that of Lemma 2 in Cai et al. (2011) and omitted.

Lemma 3.6.3. *Under Assumptions 3.1.1, D and 3.2.2-3.3.2, as* $n \to \infty$*,*

$$\frac{V(1/\beta)}{U(1/p)} - c^{\gamma} = O\left(\max\{|A(1/p)|, |A_R(1/p)|\}\right).$$

Proof. Write $\mathcal{Q} = V(1/\beta)S_n$ with

$$S_n := \left\{\mathbf{x} : \frac{D(V(1/\beta)\mathbf{x})}{h(V(1/\beta))} \leq \frac{\beta}{h(V(1/\beta))}\right\}$$

By Assumption 3.3.1, we know that $h(V(1/\beta))/\beta) = 1 + o(\alpha(V(1/\beta)))$, or,

$$\log(h(V(1/\beta))/\beta) = o(\alpha(V(1/\beta))); \tag{3.6.2}$$

see, e.g., Exercise 2.11 in de Haan and Ferreira (2006)). Lemma 3.6.2 then implies that, for some $\delta > 0$ and large n,

$$S_n \subset \left\{\mathbf{x} : \frac{D(V(1/\beta)\mathbf{x})}{h(V(1/\beta))} \leq e\right\} \subset \{\mathbf{z} : \|\mathbf{z}\| > \delta\}$$

It then follows from Lemma 3.6.1, in conjunction with (3.6.2), the existence of a sequence $\varepsilon_n = O(\alpha(V(1/\beta)))$ such that, for large n

$$e^{\varepsilon_n}S = \left\{\mathbf{z} : w(\mathbf{z}) \leq e^{-\xi\varepsilon_n}\right\} \subset S_n \subset \left\{\mathbf{z} : w(\mathbf{z}) \leq e^{\xi\varepsilon_n}\right\} = e^{-\varepsilon_n}S, \tag{3.6.3}$$

It follows that

$$\left|\frac{p}{\mathbb{P}(R \geq V(1/\beta))} - c\right| \leq \left|\frac{P(V(1/\beta)e^{\varepsilon_n}S)}{\mathbb{P}(R \geq V(1/\beta))} - c\right| + \left|\frac{P(V(1/\beta)e^{-\varepsilon_n}S)}{\mathbb{P}(R \geq V(1/\beta))} - c\right|,$$

where, for either choice of the sign,

$$\begin{aligned}\left|\frac{P(V(1/\beta)e^{\pm\varepsilon_n}S)}{\mathbb{P}(R \geq V(1/\beta))} - c\right| &\leq \left|\frac{P(V(1/\beta)e^{\pm\varepsilon_n}S)}{\mathbb{P}(R \geq V(1/\beta)e^{\pm\varepsilon_n})} - c\right| \frac{\mathbb{P}(R \geq V(1/\beta)e^{\pm\varepsilon_n})}{\mathbb{P}(R \geq V(1/\beta))} \\ &\quad + c\left|\frac{\mathbb{P}(R \geq V(1/\beta)e^{\pm\varepsilon_n})}{\mathbb{P}(R \geq V(1/\beta))} - e^{\mp\varepsilon_n/\gamma}\right| + c|e^{\mp\varepsilon_n/\gamma} - 1| \\ &= O(\alpha(V(1/\beta))) + O(\alpha(V(1/\beta))) + O(\varepsilon_n) = O(\alpha(V(1/\beta))).\end{aligned}$$

Therefore, $\frac{p}{\mathbb{P}(R \geq V(1/\beta))} - c = O(\alpha(V(1/\beta)))$. Now, together with Theorem 2.3.9 in de Haan and Ferreira (2006), we have

$$\begin{aligned}\frac{V(1/\beta)}{U(1/p)} - c^\gamma &= \frac{U(1/P(R \geq V(1/\beta)))}{U(1/p)}(1 + O(\alpha_R(V(1/\beta)))) - c^\gamma \\ &= \left(\frac{p}{\mathbb{P}(R \geq V(1/\beta))}\right)^\gamma - c^\gamma + O(A_R(1/p)) + O(\alpha_R(V(1/\beta))) \\ &= O\left(\max\left\{|\alpha(V(1/\beta))|, |A_R(1/p)|\right\}\right).\end{aligned}$$

The lemma follows by replacing $V(1/\beta)$ by $U(1/p)$ is the last line. □

Lemma 3.6.4. *Under Assumptions D and 3.1.1-3.3.2, there exists a sequence* $\varepsilon_n = O\left(\max\left\{A(1/p), A_R(n/k)\right\}\right)$ *such that* $e^{\varepsilon_n}\widetilde{\mathcal{Q}} \subset \mathcal{Q} \subset e^{-\varepsilon_n}\widetilde{\mathcal{Q}}$ *for large* n.

If further provided Assumption 3.4.1 and 3.4.2, there also exists a sequence $\varepsilon_{n1} = O\left(\max\left\{A(1/p), A_R(n/k_1)\right\}\right)$ *such that* $e^{\varepsilon_{n1}}\widetilde{\mathcal{Q}}_1 \subset \mathcal{Q} \subset e^{-\varepsilon_{n1}}\widetilde{\mathcal{Q}}_1$ *for large* n.

Proof. We only prove the first part; the proof for the second part is completely analogous. Recall from (3.6.3) there exists a sequence $\varepsilon_n = O(\alpha(V(1/\beta))) = O(A(1/p))$ such that, for large n,

$$e^{\varepsilon_n}V(1/\beta)S \subset \mathcal{Q} \subset e^{-\varepsilon_n}V(1/\beta)S, \tag{3.6.4}$$

Further, by definition, $\widetilde{\mathcal{Q}} = U(n/k)(kc/(np))^\gamma S$ and we have

$$\begin{aligned}\sup_{\mathbf{u}\in\Theta}\left|\log\frac{U(n/k)(kc/(np))^\gamma}{V(1/\beta)}\right| &\leq \left|\log\frac{V(1/\beta)}{U(1/p)} - \log\left(c^\gamma\right)\right| + \left|\log\left(\left(\frac{k}{np}\right)^\gamma\frac{U(n/k)}{U(1/p)}\right)\right| \\ &= O\left(\max\left\{|A(1/p)|, |A_R(n/k)|\right\}\right).\end{aligned}$$

by Lemma 3.6.3 here (noting that $np = o(k)$) and Theorem 2.3.9 in de Haan and Ferreira (2006). Coupling this with (3.6.4), the claim follows. □

Lemma 3.6.5. *Under the conditions of Theorem 3.3.1, as* $n \to \infty$,

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\log\widehat{c} - \log c\right) \xrightarrow{\mathbb{P}^*} 0.$$

Proof. Denote $m = m(n) := \frac{\sqrt{k}}{\log(k/(np))}$ and write $\widehat{c} = \frac{n}{k} P_n(U(n/k)\widetilde{S}_n)$ with $\widetilde{S}_n := \frac{R_{n-k:n}}{U(n/k)} \widehat{S}$. Applying the delta method (with the logarithm function), condition (3.3.2) yields that

$$m \cdot \sup_{\mathbf{u} \in \Theta} \left| \log \widehat{\rho}_S(\mathbf{u}) - \log \left(w(\mathbf{u})\right)^{\xi} \right| \xrightarrow{\mathbb{P}^*} 0.$$

On the other hand, Assumption 3.2.2 implies that $\frac{R_{n-k:n}}{U(n/k)} - 1 = O_{\mathbb{P}}(k^{-1/2})$; see, e.g., Theorem 2.4.8 in de Haan and Ferreira (2006). Applying the delta method yields that $\log \frac{R_{n-k:n}}{U(n/k)} = O_{\mathbb{P}}(k^{-1/2}) = o_{\mathbb{P}}(m^{-1})$.

Hence, for all $\varepsilon > 0$,

$$\mathbb{P}_* \left(e^{\varepsilon/m} S \subset \widetilde{S}_n \subset e^{-\varepsilon/m} S \right) \to 1. \tag{3.6.5}$$

Using Chebyshev's inequality we can show that, with either choice of sign,

$$\left| \frac{n}{k} P_n \left(U\left(\frac{n}{k}\right) e^{\pm\varepsilon/m} S \right) - \frac{n}{k} P \left(U\left(\frac{n}{k}\right) \pm e^{\varepsilon/m} S \right) \right| = o_{\mathbb{P}} \left(\frac{\log(k/(np))}{\sqrt{k}} \right).$$

Moreover, for all $\varepsilon > 0$ and large n, with either choice of sign,

$$\begin{aligned}
&\left| \frac{n}{k} P \left(U\left(\frac{n}{k}\right) e^{\pm\varepsilon/m} S \right) - c \right| \\
\leq &\left| \frac{P(U(n/k) e^{\pm\varepsilon/m} S)}{\mathbb{P}(R \geq U(n/k) e^{\pm\varepsilon/m})} - c \right| \frac{n}{k} \mathbb{P} \left(R \geq U\left(\frac{n}{k}\right) e^{\pm\varepsilon/m} \right) \\
&+ c \left| \frac{n}{k} \mathbb{P} \left(R \geq U\left(\frac{n}{k}\right) e^{\pm\varepsilon/m} \right) - (e^{\pm\varepsilon/m})^{-1/\gamma} \right| + c \left| (e^{\pm\varepsilon/m})^{-1/\gamma} - 1 \right| \\
= &O(\alpha\left(U(n/k)\right)) + O(A(n/k)) + c \left| (e^{\pm\varepsilon/m})^{-1/\gamma} - 1 \right| = O(\varepsilon m^{-1}),
\end{aligned}$$

from Assumptions 3.3.1 and 3.3.2. Hence, with (3.6.5) we know with inner probability tending to 1,

$$|\widehat{c} - c| \leq \left| \frac{n}{k} P_n \left(U\left(\frac{n}{k}\right) e^{\varepsilon/m} S \right) - c \right| + \left| \frac{n}{k} P_n \left(U\left(\frac{n}{k}\right) e^{-\varepsilon/m} S \right) - c \right| = O_{\mathbb{P}} \left(\frac{\varepsilon}{m} \right).$$

Since $\varepsilon > 0$ is arbitrary, we have $m\,|\widehat{c} - c| = o_{\mathbb{P}}(1)$ and the claim follows with an application of the delta method. $\square$

Proof of Theorem 3.3.1. We shall only show the first part of (3.3.3); the second part and (3.3.4) can be shown analogously, and the measurability is shown in the supplementary document (Lemma 3.8.1). Note that (3.2.4) implies that

$$\mathbb{P}_*(\Omega_n) := \mathbb{P}_*\left(0 < \frac{1}{2}\inf_{\mathbf{u}\in\Theta}(w(\mathbf{u}))^\xi < \widehat{\rho}_S(\mathbf{u}) < 2\sup_{\mathbf{u}\in\Theta}(w(\mathbf{u}))^\xi < \infty\right) \to 1$$

On Ω_n for large n, there then exists $\varepsilon_n, M > 0$ such that $0 < \varepsilon_n \le M_n < \infty$ and $M_n\mathbb{C} \subset \mathcal{Q}, \widetilde{\mathcal{Q}}, \widehat{\mathcal{Q}} \subset \varepsilon_n\mathbb{C}$; see also Assumption D. It follows that all the pairwise (directed) logarithm distances between $\mathcal{Q}$, $\widetilde{\mathcal{Q}}$ and $\widehat{\mathcal{Q}}$ are finite and

$$-\Delta_1(\widetilde{\mathcal{Q}}, \mathcal{Q}) \le \Delta_1(\mathcal{Q}, \widehat{\mathcal{Q}}) - \Delta_1(\widetilde{\mathcal{Q}}, \widehat{\mathcal{Q}}) \le \Delta_1(\mathcal{Q}, \widetilde{\mathcal{Q}}). \tag{3.6.6}$$

Moreover, since $\Delta_1(\mathcal{Q}, \widetilde{\mathcal{Q}}) \le \delta \Leftrightarrow \mathcal{Q} \subset e^{-\delta}\widetilde{\mathcal{Q}}$ and $\Delta_1(\widetilde{\mathcal{Q}}, \mathcal{Q}) \le \delta \Leftrightarrow \widetilde{\mathcal{Q}} \subset e^{-\delta}\mathcal{Q}$ for all $\delta > 0$, by Lemma 3.6.4 and assumption we have

$$\left|\Delta_1(\widetilde{\mathcal{Q}}, \mathcal{Q})\right| + \left|\Delta_1(\mathcal{Q}, \widetilde{\mathcal{Q}})\right| = O(\max\{|A(1/p)|, |A_R(n/k)|\}) = o\left(\frac{\log(k/(np))}{\sqrt{k}}\right).$$

Combining this with (3.6.6) yields

$$\frac{\sqrt{k}}{\log(k/(\log(np)))}\left|\Delta_1(\mathcal{Q}, \widehat{\mathcal{Q}}) - \Delta_1(\widetilde{\mathcal{Q}}, \widehat{\mathcal{Q}})\right| = o_{\mathbb{P}}(1),$$

and, therefore, it suffices to prove the claim with $\widetilde{\mathcal{Q}}$ replacing $\mathcal{Q}$, that is,

$$\begin{aligned}
&\left|\frac{\sqrt{k}}{\log(k/(\log(np)))}\Delta_1(\widetilde{\mathcal{Q}}, \widehat{\mathcal{Q}}) - \Gamma_n\right| \\
&= \left|\frac{\sqrt{k}}{\log(k/(np))}\sup_{\mathbf{u}\in\Theta}\log\frac{R_{n-k:n}(k\widehat{c}/(np))^{\widehat{\gamma}}\widehat{\rho}_S(\mathbf{u})}{U(n/k)(kc/(np))^{\gamma}(w(\mathbf{u}))^\xi} - \Gamma_n\right| \\
&\le \frac{1}{|\log(k/(np)|}\left|\sqrt{k}\log\frac{R_{n-k:n}}{U(n/k)}\right| + \left|\frac{\log c}{\log(k/(np))}\right||\Gamma_n| + \widehat{\gamma}\left|\frac{\sqrt{k}}{\log(k/(np))}\log\frac{\widehat{c}}{c}\right| \\
&+ \sup_{\mathbf{u}\in\Theta}\left|\frac{\sqrt{k}}{\log(k/(np))}\left(\widehat{\rho}_S(\mathbf{u}) - (w(\mathbf{u}))^\xi\right)\right| \xrightarrow{\mathbb{P}^*} 0,
\end{aligned}$$

where in the last two lines the convergence of the first term follows by $\sqrt{k}\log\frac{R_{n-k:n}}{U(n/k)} = O_{\mathbb{P}}(1)$, the second term by the well-known result in univariate extreme value theory that $\Gamma_n \xrightarrow{d} N(0, \gamma^2)$ (see, e.g., Theorem 3.2.5 in de Haan and Ferreira, 2006), the third term by Lemma 3.6.5 and the consistency of $\widehat{\gamma}$, and the last term by assumption. □

Proof of Theorem 3.2.1. It suffices to show that $\Delta_2(\mathcal{Q}, \widehat{\mathcal{Q}}) \xrightarrow{\mathbb{P}^*} 0$. The proof is then very similar to that of Theorem 3.3.1 above, allowing the convergence rates bounded by $A(1/p)$ and $A_R(n/k)$ to be $o(1)$ instead and using the consistency results $R_{n-k:n}/U(n/k) \xrightarrow{\mathbb{P}} 1$, $\widehat{\gamma} \xrightarrow{\mathbb{P}} \gamma$ and $\widehat{c} \xrightarrow{\mathbb{P}^*} c$; see, e.g., Theorem 2.4.8 and 3.2.2 in de Haan and Ferreira (2006) for the first two. The proof of $\widehat{c} \xrightarrow{\mathbb{P}^*} c$ is completely analogous to that of Lemma 3.6.5 by replacing $m = m(n) = 1$ therein and noticing $\frac{\log(k/(np))}{\sqrt{k}} = \frac{\log(k)}{\sqrt{k}} - \frac{\log(np)}{\sqrt{k}} \to 0$. The measurability is shown in the supplementary document (Lemma 3.8.1). $\square$

We denote the class of half-spaces and complements of centered balls that bounded away from the origin with distances larger than $\delta > 0$ by

$$\bar{\mathcal{H}}_\delta^+ := \{H_{r,\mathbf{u}} \in \mathcal{H} : r > \delta, \mathbf{u} \in \Theta\} \cup \{r\mathbb{C} : r > \delta\}.$$

Clearly, it forms a Vapnik-Chervonenkis class (VC-class) and

$$\sup_{B \in \bar{\mathcal{H}}_\delta^+} \nu(B) = \nu\left(\cup_{B \in \bar{\mathcal{H}}_\delta^+} B\right) = \nu(\delta\mathbb{C}) = \delta^{-1/\gamma} < \infty.$$

To prove Theorems 3.4.1 and 3.4.2 we introduce a *pseudo* and a true tail empirical process indexed by the elements of $\bar{\mathcal{H}}_0^+ = \bigcup_{\delta > 0} \bar{\mathcal{H}}_\delta^+$ given by

$$\widetilde{v}_{n,k_1}(\cdot) := \sqrt{k_1}\left(\widetilde{\nu}_{n1}(\cdot) - \nu(\cdot)\right), \; v_{n,k_1}(\cdot) := \sqrt{k_1}\left(\widehat{\nu}_1(\cdot) - \nu(\cdot)\right)$$

where $\widetilde{\nu}_{n1}$ and $\widehat{\nu}_1$ are a pseudo and a true estimator of ν given by

$$\widetilde{\nu}_{n1}(B) = \frac{n}{k_1} P_n\left(U\left(\frac{n}{k_1}\right) B\right), \; \widehat{\nu}_1(B) = \frac{n}{k_1} P_n\left(R_{n-k_1:n} B\right) \quad B \in \bar{\mathcal{H}}_0^+.$$

Analogously also define $\widetilde{v}_{n,k}$ and $v_{n,k}$ (and $\widetilde{\nu}_n$) on $\bar{\mathcal{H}}^+$. Denote '$\xrightarrow{w}$' as the weak convergence in the 'classical' sense; see, e.g., (34) in Gänssler (1983), page 65.

Lemma 3.6.6. *Under the conditions of Theorem 3.4.1 with $\psi > 0$, for all $\delta \in (0, 1)$, as $n \to \infty$,*

$$\left\{\widetilde{v}_{n,k_1}(B_1), \widetilde{v}_{n,k}(B_2), B_1, B_2 \in \bar{\mathcal{H}}_\delta^+\right\} \xrightarrow{w} \left\{W_{\nu,1}(B_1), W_\nu(B_2), B_1, B_2 \in \bar{\mathcal{H}}_\delta^+\right\},$$

where $W_{\nu,1}$ is a ν-Brownian motion, that is, a mean-zero tight Gaussian process with the covariance structure

$$Cov(W_{\nu,1}(B_1), W_{\nu,1}(B_2)) = \nu(B_1 \cap B_2) \quad B_1, B_2 \in \bar{\mathcal{H}}_\delta^+,$$

which has bounded and uniformly d_ν-continuous sample paths and d_ν is a semimetric defined on $\bar{\mathcal{H}}_0^+$ given by

$$d_\nu(B_1, B_2) = \nu(B_1 \Delta B_2) \quad B_1, B_2 \in \bar{\mathcal{H}}_0^+,$$

and W_ν is an independent copy of $W_{\nu,1}$.

Proof. Let us postpone the measurability issue to the end and first show that, in the product space of bounded functions on $\bar{\mathcal{H}}_\delta^+$,

$$\begin{aligned} &\left\{\widetilde{v}_{n,k_1}(B_1) - \mathbb{E}\left(\widetilde{v}_{n,k_1}(B_1)\right), \widetilde{v}_{n,k}(B_2) - \mathbb{E}\left(\widetilde{v}_{n,k}(B_2)\right), B_1, B_2 \in \bar{\mathcal{H}}_\delta^+\right\} \\ &\quad \rightsquigarrow \left\{W_{\nu,1}(B_1), W_\nu(B_2), B_1, B_2 \in \bar{\mathcal{H}}_\delta^+\right\}, \end{aligned} \tag{3.6.7}$$

where '$\rightsquigarrow$' denotes the general notion of weak convergence introduced by Hoffmann-Jørgensen (1991). The convergence in the first coordinate alone (and analogously also in the second coordinate) is readily available by applying Theorem 3 in Einmahl (1997), which requires verification of that

$$\sup_{B \in \mathcal{A}_\delta} \left| \frac{n}{k_1} P\left(U\left(\frac{n}{k_1}\right) B\right) - \nu(B) \right| \to 0 \tag{3.6.8}$$

where $\mathcal{A}_\delta = \{B \cap B' : B, B' \in \bar{\mathcal{H}}_\delta^+\}$. Showing (3.6.8) is very similar to that for Lemma 1 in Einmahl et al. (2015a) and its details are provided in our supplementary document (Lemma 3.8.4). Since the joint (asymptotic) tightness is a consequence of the coordinate-wise (asymptotic) tightness, it remains to verify that the finite-dimensional distributions of our processes converge weakly; see, e.g., exercise 1.5.3 in van der Vaart and Wellner (1996). While it is easy to show that the linear combination of the components converges using the univariate Lindeberg-Feller central limit theorem (noting that $k_1/k \to 0$), the multivariate weak convergence follows.

Moreover, cf. Lemma 1, we can show that

$$\sup_{B\in\bar{\mathcal{H}}_\delta^+} |\mathbb{E}(\widetilde{v}_{n,k_1}(B))| = \sqrt{k_1} \sup_{B\in\bar{\mathcal{H}}_\delta^+} \left| \frac{n}{k_1} P\left(U\left(\frac{n}{k_1}\right)B\right) - \nu(B)\right|$$
$$= O(\sqrt{k_1}A(n/k_1)) \to 0,$$

and similarly $\sup_{B\in\bar{\mathcal{H}}_\delta^+} |\mathbb{E}(v_{n,k}(B))| = O(\sqrt{k}A(n/k)) \to 0$. Hence,

$$(\widetilde{v}_{n,k_1}, \widetilde{v}_{n,k}) \rightsquigarrow (W_{\nu,1}, W_\nu), \quad \text{on } \bar{\mathcal{H}}_\delta^+ \times \bar{\mathcal{H}}_\delta^+.$$

It remains to check the measurability, which follows from the so-called *universal separability* of index class $\bar{\mathcal{H}}_\delta^+$ (see Lemma 3.8.2 in our supplementary document), i.e. condition (SE) in Gänssler (1983), page 108. For details see pages 70 and 105-107 in Gänssler (1983), taking the measure μ there as our exponent measure ν. □

Lemma 3.6.7. *Under the conditions of Theorem 3.4.1 with $\psi > 0$, for all $\delta \in (0,1)$, under a Skorohod-Dudley-Wichura construction (see, e.g., Gänssler, 1983, page 82), as $n \to \infty$,*

$$\sqrt{k_1}\log\frac{R_{n-k_1:n}}{U(n/k_1)} \xrightarrow{a.s.} \gamma W_{\nu,1}(\mathbb{C}), \quad \sqrt{k}\log\frac{R_{n-k:n}}{U(n/k)} \xrightarrow{a.s.} \gamma W_\nu(\mathbb{C}),$$
$$\sup_{B\in\bar{\mathcal{H}}_\delta^+} |v_{n,k_1}(B) - \{W_{\nu,1}(B) - \nu(B)W_{\nu,1}(\mathbb{C})\}| \xrightarrow{a.s.} 0,$$
$$\sup_{B\in\bar{\mathcal{H}}_\delta^+} |v_{n,k}(B) - \{W_\nu(B) - \nu(B)W_\nu(\mathbb{C})\}| \xrightarrow{a.s.} 0,$$

and, therefore, $\Gamma_n \xrightarrow{\mathbb{P}} \int_1^\infty W_\nu(s\mathbb{C})\frac{ds}{s} - \gamma W_\nu(\mathbb{C}) =: \Gamma \sim N(0,\gamma^2)$. Note that all the random elements here only equal in distribution to the original ones.

Proof. With Lemma 3.6.6, invoking a Skorohod-Dudley-Wichura construction, we can start from

$$\sup_{B\in\bar{\mathcal{H}}_{\delta/2}^+} |\widetilde{v}_{n,k_1}(B) - W_{\nu,1}(B)| \xrightarrow{a.s.} 0, \quad \sup_{B\in\bar{\mathcal{H}}_{\delta/2}^+} |\widetilde{v}_{n,k}(B) - W_\nu(B)| \xrightarrow{a.s.} 0 \quad (3.6.9)$$

where all processes are defined on the same probability space and only equal in distribution to the original ones. In the following we only prove the statements for k_1; the proof for those with k is completely analogous. The first statement in (3.6.9) implies that

$$\sup_{x\geq\delta/2}\left|\sqrt{k_1}\left\{\frac{n}{k_1}P_n\left(xU\left(\frac{n}{k_1}\right)\mathbb{C}\right)-x^{-1/\gamma}\right\}-W_{\nu,1}(x\mathbb{C})\right|\xrightarrow{a.s.}0,$$

Applying Vervaat (1972) inversion lemma around $x=1$ yields that

$$\sqrt{k_1}\left(\frac{R_{n-k_1:n}}{U(n/k_1)}-1\right)\xrightarrow{a.s.}\gamma W_{\nu,1}\left(\mathbb{C}\right), \tag{3.6.10}$$

and then the first statement in the lemma follows by the delta method. Hence,

$$\begin{aligned}
&\sup_{B\in\bar{\mathcal{H}}_\delta}\left|v_{n,k_1}\left(B\right)-\left(W_{\nu,1}(B)-\nu(B)W_{\nu,1}(\mathbb{C})\right)\right|\\
&\leq\sup_{B\in\bar{\mathcal{H}}_\delta}\left|\widetilde{v}_{n,k_1}\left(\frac{R_{n-k_1:n}}{U(n/k_1)}B\right)-W_{\nu,1}\left(\frac{R_{n-k_1:n}}{U(n/k_1)}B\right)\right|\\
&\quad+\sup_{B\in\bar{\mathcal{H}}_\delta}\nu(B)\left|\sqrt{k_1}\left\{\left(\frac{R_{n-k_1:n}}{U(n/k_1)}\right)^{-1/\gamma}-1\right\}+W_{\nu,1}(\mathbb{C})\right|\\
&\quad+\sup_{B\in\bar{\mathcal{H}}_\delta}\left|W_{\nu,1}\left(\frac{R_{n-k_1:n}}{U(n/k_1)}B\right)-W_{\nu,1}(B)\right|\xrightarrow{a.s.}0+0+0=0,
\end{aligned}$$

where the convergence of the first term follows by (3.6.9) and $\frac{R_{n-k_1:n}}{U(n/k_1)}\xrightarrow{a.s.}1$, the second term by applying the delta method with (3.6.10) and the last term by the uniform d_ν-continuity of W_ν.

It remains to show $\Gamma_n=\int_1^\infty v_{n,k}\left(s\mathbb{C}\right)\frac{ds}{s}\xrightarrow{\mathbb{P}}\Gamma$, which is a standard exercise in extreme value theory; see, e.g., Example 5.1.5 in de Haan and Ferreira (2006). Note that for this we need the Chibisov-O'Reilly theorem (see, e.g., Shorack and Wellner 1986, p. 462) and Lebesgue dominated convergence theorem, and we refer the details to the aforementioned book chapters. □

Lemma 3.6.8. *Suppose Assumption R holds. For all* $\mathbf{x}\in\mathbb{R}^d\setminus\{\mathbf{0}\}$,

$$HD(\mathbf{x};\nu)=\inf_{\mathbf{u}\in\Theta_0}\nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}}).$$

Proof. Lemma 2.5.4 states

$$HD(\mathbf{x};\nu) = \inf_{\mathbf{u}\in\Theta,\mathbf{u}^T\mathbf{x}\geq\delta_0^\gamma\|\mathbf{x}\|} \nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}})$$

We further know from Lemma 2.5.3 that $\nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}}) = (\mathbf{u}^T\mathbf{x})^{-1/\gamma}\nu(H_{1,\mathbf{u}})$ is continuous in $\{\mathbf{u}\in\Theta : \mathbf{u}^T\mathbf{x}\geq\delta_0^\gamma\|\mathbf{x}\|\}\subset\Theta$. It follows that $HD(\mathbf{x};\nu)\geq \inf_{\mathbf{u}\in\Theta_0}\nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}})$. The lemma follows since we also have

$$HD(\mathbf{x};\nu)\leq\inf_{\mathbf{u}\in\Theta_0}\nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}})$$

by definition. □

Lemma 3.6.9. *Under the conditions of Theorem 3.4.1, as $n\to\infty$,*

$$\frac{\sqrt{k}}{\log(k_1/(np))}(\log\widehat{c}_1 - \log c)\xrightarrow{\mathbb{P}} 0.$$

Proof. Write $\widehat{c}_1 = \frac{n}{k}P_n(U(n/k)\widetilde{S}_{n1})$ with $\widetilde{S}_{n1} = \left\{\mathbf{x}\in\mathbb{R}^d : \widehat{HD}(\mathbf{x};\widetilde{\nu}_n)\leq 1\right\}$.

We can take some $\delta\in(0,\delta_0^\gamma)$ such that $\mathbb{P}\left(\widetilde{S}_{n1}\subset\delta\mathbb{C}\right)\to 1$ since

$$\inf_{\|\mathbf{x}\|\leq\delta}\widehat{HD}(\mathbf{x};\widetilde{\nu}_n)\geq\inf_{\mathbf{u}\in\Theta_0}\widetilde{\nu}_n\left(\delta H_{1,\mathbf{u}}\right)\xrightarrow{\mathbb{P}}\inf_{\mathbf{u}\in\Theta_0}\nu\left(\delta H_{1,\mathbf{u}}\right) = \delta_0/\delta^{1/\gamma} > 1.$$

Moreover, with probability tending to 1

$$\widehat{HD}(\mathbf{x};\widetilde{\nu}_n) = \inf_{\mathbf{u}^T\mathbf{x}\geq\delta_0^\gamma\delta,\mathbf{u}\in\Theta_0}\widetilde{\nu}_n(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}}) \text{ for all } \|\mathbf{x}\|\geq\delta \tag{3.6.11}$$

since, for any $c\in((1-\delta_0)^\gamma,1)$

$$\begin{aligned}
&\inf_{\|\mathbf{x}\|\geq\delta}\inf_{\mathbf{u}^T\mathbf{x}<\delta_0^\gamma\delta,\mathbf{u}\in\Theta_0}\widetilde{\nu}_n(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}}) - \sup_{\|\mathbf{x}\|\geq\delta}\widehat{HD}(\mathbf{x};\widetilde{\nu}_n)\\
&\geq\inf_{\mathbf{u}\in\Theta_0}\widetilde{\nu}_n(H_{\delta_0^\gamma\delta,\mathbf{u}}) - \sup_{\mathbf{u}\in\Theta_0}\widetilde{\nu}_n(H_{c\delta,\mathbf{u}})\xrightarrow{\mathbb{P}}\inf_{\mathbf{u}\in\Theta_0}\nu(H_{\delta_0^\gamma\delta,\mathbf{u}}) - \sup_{\mathbf{u}\in\Theta_0}\nu(H_{c\delta,\mathbf{u}})\\
&= (\delta_0^\gamma\delta)^{-1/\gamma}\cdot\delta_0 - (c\delta)^{-1/\gamma}\sup_{\mathbf{u}\in\Theta_0}\nu(H_{1,\mathbf{u}}) > \delta^{-1/\gamma}(1-c^{-1/\gamma}(1-\delta_0)) > 0.
\end{aligned}$$

Similarly, even easier, with Lemma 3.6.8 we can also show

$$HD(\mathbf{x};\nu) = \inf_{\mathbf{u}^T\mathbf{x}\geq\delta_0^\gamma,\mathbf{u}\in\Theta_0}\nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}}) \text{ for all } \|\mathbf{x}\|\geq\delta. \tag{3.6.12}$$

Hence, using (3.6.11) and (3.6.12), we know with probability tending to 1

$$\sup_{\mathbf{x}\geq\delta}\left|\widehat{HD}(\mathbf{x};\widetilde{\nu}_n)-HD(\mathbf{x};\nu)\right|\leq\sup_{r\geq\delta_0^\gamma\delta,\mathbf{u}\in\Theta}|\widetilde{\nu}_n(H_{r,\mathbf{u}})-\nu(H_{r,\mathbf{u}})|=O_{\mathbb{P}}(k^{-1/2}).$$

It follows that, similar as (3.6.5), for all $\varepsilon>0$ we have that

$$\mathbb{P}\left(e^{\varepsilon/m}S\subset\widetilde{S}_{n1}\subset e^{-\varepsilon/m}S\right)\to 1,$$

where $m=m(n)=\sqrt{k}/\log(k_1/(np))$. The rest is then completely analogous to the final part (after equation (3.6.5)) of the proof of Lemma 3.6.5. □

Proof of Theorem 3.4.1. We only prove the case $\psi>0$ here. The proof for $\psi=0$ is very similar to that of Theorem 3.3.1 and hence omitted. We only prove the claim for $\bar{C}_\tau^+$, that is,

$$\liminf_{n\to\infty}\mathbb{P}(\bar{C}_\tau^+)=\liminf_{n\to\infty}\mathbb{P}(\Delta_1(\mathcal{Q},\widehat{\mathcal{Q}}_1)\leq q_1(\tau))\geq\tau;$$

the proof of those for $\bar{C}_\tau^-$ and $\bar{C}_\tau$ are analogous. With Lemma 3.6.4 (to apply which we verified Assumption 3.3.2 in the supplementary document, Lemma 3.8.5), cf. the proof of Theorem 3.3.1, we only need to prove the claim with $\mathcal{Q}$ replaced by

$$\widetilde{\mathcal{Q}}_1:=U\left(\frac{n}{k_1}\right)\left(\frac{k_1c}{np}\right)^\gamma S=\bigcup_{\mathbf{u}\in\Theta_0}\left\{\mathbf{x}\in\mathbb{R}^d:\mathbf{u}^T\mathbf{x}>g(\mathbf{u})\right\},$$

where $g(\cdot):=U(n/k_1)(k_1c/(np))^\gamma(\nu(H_{1,\cdot}))^\gamma$. We can also write

$$\widehat{\mathcal{Q}}_1=R_{n-k_1:n}\left(\frac{k_1\widehat{c}_1}{np}\right)^{\widehat{\gamma}}\widehat{S}_1=\bigcup_{\mathbf{u}\in\Theta_0}\left\{\mathbf{x}\in\mathbb{R}^d:\mathbf{u}^T\mathbf{x}>\widehat{g}(\mathbf{u})\right\}.$$

where $\widehat{g}(\cdot):=R_{n-k_1:n}(k_1\widehat{c}_1/(np))^{\widehat{\gamma}}\,(\widehat{\nu}_1(H_{1,\cdot}))^{\widehat{\gamma}}$. Now, for all $\mathbf{u}\in\Theta$,

$$\begin{aligned}\sqrt{k_1}\psi_n\log\frac{\widehat{g}(\mathbf{u})}{g(\mathbf{u})}=&\Gamma_n+\psi_n\sqrt{k_1}\log\frac{R_{n-k_1:n}}{U(n/k)}+\psi_n\sqrt{k_1}\widehat{\gamma}\log\frac{\widehat{\nu}_1(H_{1,\mathbf{u}})}{\nu(H_{1,\mathbf{u}})}\\&+\widehat{\gamma}\frac{\sqrt{k}}{\log(k_1/(np))}\log\frac{\widehat{c}_1}{c}+\frac{\log c+\log\nu(H_{1,\mathbf{u}})}{\log(k_1/(np))}\cdot\Gamma_n,\end{aligned}$$

where the last two terms converge to zero in probability uniformly for $\mathbf{u} \in \Theta$ by Lemma 3.6.9 and the facts that $\Gamma_n = O_{\mathbb{P}}(1)$ and $np = o(k_1)$. Hence, applying the delta method with Lemma 3.6.7 yields that

$$\left\{ \sqrt{k_1}\psi_n \log \frac{\widehat{g}(\mathbf{u})}{g(\mathbf{u})} : \mathbf{u} \in \Theta_0 \right\} \xrightarrow{w} \{\gamma \mathbb{Z}(\mathbf{u}) : \mathbf{u} \in \Theta_0\},$$

where $\mathbb{Z}$ is a mean-zero tight Gaussian process on Θ given by

$$\mathbb{Z}(\mathbf{u}) := \frac{1}{\gamma}\Gamma + \psi \frac{W_{\nu,1}(H_{1,\mathbf{u}})}{\nu(H_{1,\mathbf{u}})}, \quad \mathbf{u} \in \Theta, \tag{3.6.13}$$

with covariance structure

$$\begin{aligned} \operatorname{Cov}(\mathbb{Z}(\mathbf{u}), \mathbb{Z}(\mathbf{v})) &= \frac{1}{\gamma^2}\operatorname{Var}(\Gamma) + \psi^2 \frac{\operatorname{Cov}(W_\nu(H_{1,\mathbf{u}}), W_\nu(H_{1,\mathbf{v}}))}{\nu(H_{1,\mathbf{u}})\nu(H_{1,\mathbf{v}})} \\ &= 1 + \psi^2 \frac{\nu(H_{1,\mathbf{u}} \cap H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{u}})\nu(H_{1,\mathbf{v}})}, \quad \mathbf{u}, \mathbf{v} \in \Theta. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}(\Delta_1(\widetilde{\mathcal{Q}}_1, \widehat{\mathcal{Q}}_1) \le q_1(\tau)) &\ge \mathbb{P}\left(\sqrt{k_1}\psi_n \sup_{\mathbf{u} \in \Theta_0} \log \frac{\widehat{g}(\mathbf{u})}{g(\mathbf{u})} \le q_1(\tau) \right) \\ \to \mathbb{P}\left(\sup_{\mathbf{u} \in \Theta_0} \mathbb{Z}(\mathbf{u}) \le q_1(\tau) \right) &= \mathbb{P}\left(\sup_{\mathbf{u} \in \Theta} \mathbb{Z}(\mathbf{u}) \le q_1(\tau) \right) = \tau, \end{aligned}$$

where the penultimate equality results from the continuity of $\mathbb{Z}$. □

Proof of Theorem 3.4.2. We only prove the case $\psi > 0$ here. The proof for $\psi = 0$ is very similar to that of Theorem 3.3.1 and hence omitted. With Lemma 3.6.4 (to apply which we verified Assumption 3.3.2 in the supplementary document, Lemma 3.8.5), cf. the proof of Theorem 3.3.1, it suffices to prove the theorem with $\mathcal{Q}$ replaced by $\widetilde{\mathcal{Q}}_1 := U\left(\frac{n}{k_1}\right)\left(\frac{k_1 c}{np}\right)^\gamma S$. By definition we have

$$\widetilde{\mathcal{Q}}_1 =: \{r\mathbf{u} : r > \widetilde{\rho}_1(\mathbf{u}), \mathbf{u} \in \Theta\}, \;\; \widehat{\mathcal{Q}}_1 = \{r\mathbf{u} : r > \widehat{\rho}_1(\mathbf{u}), \mathbf{u} \in \Theta\}$$

and

$$\sqrt{k_1}\psi_n \log \frac{\widehat{\rho}_1(\mathbf{u})}{\widetilde{\rho}_1(\mathbf{u})}$$

$$
\begin{aligned}
&= \sqrt{k_1}\psi_n \log \frac{R_{n-k:n}(k\widehat{c}_1/(np))^{\widehat{\gamma}} HD(\mathbf{u};\widehat{\nu}_1^*)^{\widehat{\gamma}}}{U(n/k)(k\nu(S)/(np))^{\gamma} HD(\mathbf{u};\nu)^{\gamma}} \\
&= \Gamma_n + \psi_n\sqrt{k_1}\left\{\log \frac{R_{n-k_1:n}}{U(n/k_1)} + \widehat{\gamma}\log\frac{\widehat{HD}(\mathbf{u};\widehat{\nu}_1^*)}{HD(\mathbf{u};\nu)}\right\} + \frac{\widehat{\gamma}\sqrt{k}}{\log(k_1/(np))}\log\frac{\widehat{c}_1}{c} \\
&\quad + \left(\frac{\log c}{\log(k_1/(np))} + \sqrt{\frac{k_1}{k}}\log(HD(\mathbf{u};\nu))\right)\Gamma_n
\end{aligned}
$$

where the last two terms converge to zero in probability by Lemma 3.6.9 and the facts that $\Gamma_n = O_{\mathbb{P}}(1)$ and $k_1/k \to 0$. Under the Skorohod-Dudley-Wichura construction of Lemma 3.6.7 with some $\delta \in (0,1)$,

$$
\Gamma_n \xrightarrow{\mathbb{P}} \Gamma, \ \sqrt{k_1}\log\frac{R_{n-k_1:n}}{U(n/k_1)} \xrightarrow{a.s.} \gamma W_{\nu,1}(\mathbb{C}),
$$

and we shall show in the end that

$$
\sup_{\mathbf{u}\in\Theta}\left|\sqrt{k_1}\log\frac{\widehat{HD}(\mathbf{u};\widehat{\nu}_1^*)}{HD(\mathbf{u};\nu)} - \left(\inf_{\mathbf{v}\in\phi(\mathbf{u})}\frac{W_{\nu,1}(H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{v}})} - W_{\nu,1}(\mathbb{C})\right)\right| \xrightarrow{\mathbb{P}} 0. \quad (3.6.14)
$$

It follows that

$$
\left\{\sqrt{k_1}\psi_n\log\frac{\widehat{\rho}_1(\mathbf{u})}{\widetilde{\rho}_1(\mathbf{u})}, \mathbf{u}\in\Theta\right\} \xrightarrow{w} \left\{\inf_{\mathbf{v}\in\phi(\mathbf{u})}\mathbb{Z}(\mathbf{v}), \mathbf{u}\in\Theta\right\}.
$$

with $\mathbb{Z}$ defined in (3.6.13). The theorem then follows by noting that

$$
\begin{aligned}
\Delta_1(\widetilde{\mathcal{Q}}_1, \widehat{\mathcal{Q}}_1) = \sup_{\mathbf{u}\in\Theta}\log\frac{\widehat{\rho}_1(\mathbf{u})}{\widetilde{\rho}_1(\mathbf{u})},\ \Delta_1(\widehat{\mathcal{Q}}_1, \widetilde{\mathcal{Q}}_1) = \sup_{\mathbf{u}\in\Theta}\log\frac{\widetilde{\rho}_1(\mathbf{u})}{\widehat{\rho}_1(\mathbf{u})}, \\
\Delta_2(\widetilde{\mathcal{Q}}_1, \widehat{\mathcal{Q}}_1) = \sup_{\mathbf{u}\in\Theta}\left|\log\frac{\widehat{\rho}_1(\mathbf{u})}{\widetilde{\rho}_1(\mathbf{u})}\right|.
\end{aligned}
$$

It remains to verify (3.6.14), under the Skorohod-Dudley-Wichura construction above. By Lemma 3.6.7 and noting that $k_1/k \to 0$,

$$
\begin{aligned}
&\sup_{1\ge r\ge\delta_0^{\gamma},\mathbf{u}\in\Theta}\left|\sqrt{k_1}\left(\widehat{\nu}_1^*(H_{r,\mathbf{u}}) - \nu(H_{r,\mathbf{u}})\right) - \nu(H_{r,\mathbf{u}})\left(\frac{W_{\nu,1}(H_{1,\mathbf{u}})}{\nu(H_{1,\mathbf{u}})} - W_{\nu,1}(\mathbb{C})\right)\right| \\
\le &\sup_{1\ge r\ge\delta_0^{\gamma}} r^{-1/\gamma}\sup_{\mathbf{u}\in\Theta}\left|v_{n,k_1}(H_{1,\mathbf{u}}) - (W_{\nu,1}(H_{1,\mathbf{u}}) - \nu(H_{1,\mathbf{u}})W_{\nu,1}(\mathbb{C}))\right| \\
&+ \sqrt{\frac{k_1}{k}}\cdot\sqrt{k}\left|\delta_0^{-\gamma/\widehat{\gamma}} - \delta_0^{-1}\right|\sup_{\mathbf{u}\in\Theta}\widehat{\nu}_1(H_{1,\mathbf{u}}) \xrightarrow{\mathbb{P}} 0 + 0 = 0.
\end{aligned}
$$

Then following the proof of Theorem 2.1 in Arcones et al. (2006), with Assumption 3.4.3 we can show that (see Lemma 3.8.8 in our supplementary document)

$$\left\{\sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0}\widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})-HD(\mathbf{u};\nu)\right):\mathbf{u}\in\Theta\right\}$$
$$\xrightarrow{\mathbb{P}}\left\{\inf_{\mathbf{v}\in\phi(\mathbf{u})}\nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\left(\frac{W_{\nu,1}(H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{v}})}-W_{\nu,1}(\mathbb{C})\right):\mathbf{u}\in\Theta\right\},$$

Similar as (3.6.11), we can show that with probability tending to 1

$$\widehat{HD}(\mathbf{u};\widehat{\nu}_1^*)=\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0}\widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{u}})\text{ for all }\mathbf{u}\in\Theta.$$

Then substituting $\nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})$ by $HD(\mathbf{u};\nu)$ for $\mathbf{v}\in\phi(\mathbf{u})$ yields that

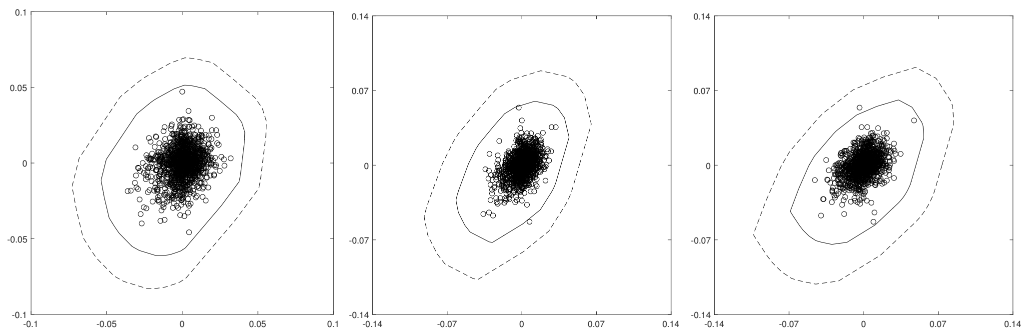
$$\left\{\sqrt{k_1}\left(\widehat{HD}(\mathbf{u};\widehat{\nu}_1^*)-HD(\mathbf{u};\nu)\right):\mathbf{u}\in\Theta\right\}$$
$$\xrightarrow{\mathbb{P}}\left\{HD(\mathbf{u};\nu)\left(\inf_{\mathbf{v}\in\phi(\mathbf{u})}\frac{W_{\nu,1}(H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{v}})}-W_{\nu,1}(\mathbb{C})\right):\mathbf{u}\in\Theta\right\}.$$

Since $0<\delta_0\leq HD(\mathbf{u};\nu)\leq 1$, the rest follows by the delta method. □

3.7 Applications to financial data

In this section, we present a real-world finance application. The dataset, downloaded from Datastream, consists of the daily international market price indexes of Standard & Poor's 500 (S&P 500) from the United States, Financial Times Stock Exchange 100 (FTSE 100) from the United Kingdom and Hang Seng from Hong Kong. The sample period is from October 15, 2010, to October 16, 2015. We use continuously compounded daily market returns, giving rise to 1305 observations for each country. Since the US market is the latest to close on any particular day, in the analysis we use the one-period ago US returns whenever the returns pair (and triple) involves S&P 500 returns. Although these returns series do not include the dividend distribution, this is not a problem for our analysis since extreme movements can hardly be generated by dividends (Poon et al., 2004).

As usual, our collected stock returns exhibit so-called *volatility clustering* behavior widely documented in the empirical finance literature: the univariate squared stock returns are moderately auto-correlated, and the Ljung-Box test rejects their serial independence at 5% level. Hence, we shall work on a filtered version of the univariate returns by calibrating a generalized autoregressive conditional heteroskedasticity GARCH(1,1) model (Bollerslev, 1986). Now, the Ljung-Box test at 5% level does not reject the serial independence of the time series of the absolute nor squared residual returns. From now on, these residual returns will also be called *filtered* returns and, for simplicity, we take the estimated model parameters in our calibrated GARCH model as the actual ones.



(a) US vs UK (b) US vs HK (c) UK vs HK

Figure 3.5: One-day ahead predicted half-space depth based quantile regions (solid) for $p = 0.001$ and the 95% asymptotically conservative confidence sets of the predicted central region (close inner region of the dashed lines) of raw returns pairs given information from Oct. 15, 2010 to Oct. 16, 2015.

Next, we check the equality of the extreme value indices for the tails of the univariate filtered returns, implied by Assumption 6. Following Cai et al. (2011) and He and Einmahl (2016), we compare the Hill estimates for the positive and negative tails of the filtered returns in all three markets for $k = 80$. In increasing order these estimates are 0.261, 0.264, 0.266, 0.276, 0.297, 0.309, between which the maximal difference is 0.048. Based on the asymptotic normality of the Hill's estimator, this gives an approximate p-value of 0.896. There is no evidence that the true extreme value indices are different.

The *predicted* quantile region of the joint raw return $\mathbf{r}_t = (r_t^{US}, r_t^{UK}, r_t^{HK})$, given the information up to time $t-1$, can be easily obtained using the *unconditional* quantile region of the filtered return $\mathbf{z}_t = (z_t^{US}, z_t^{UK}, z_t^{HK})$ by a simple affine transformation. Figure 3.5 presents the one-day ahead predicted bivariate half-space depth based quantile regions for $p = 0.001$, motivated by the recent Basel II requirement (Basel Committee on Banking Supervision, 2005), of raw returns pairs. The close inner region of the dashed lines refer to our 95% asymptotically conservative confidence set of the predicted central regions, given information in our sample period.

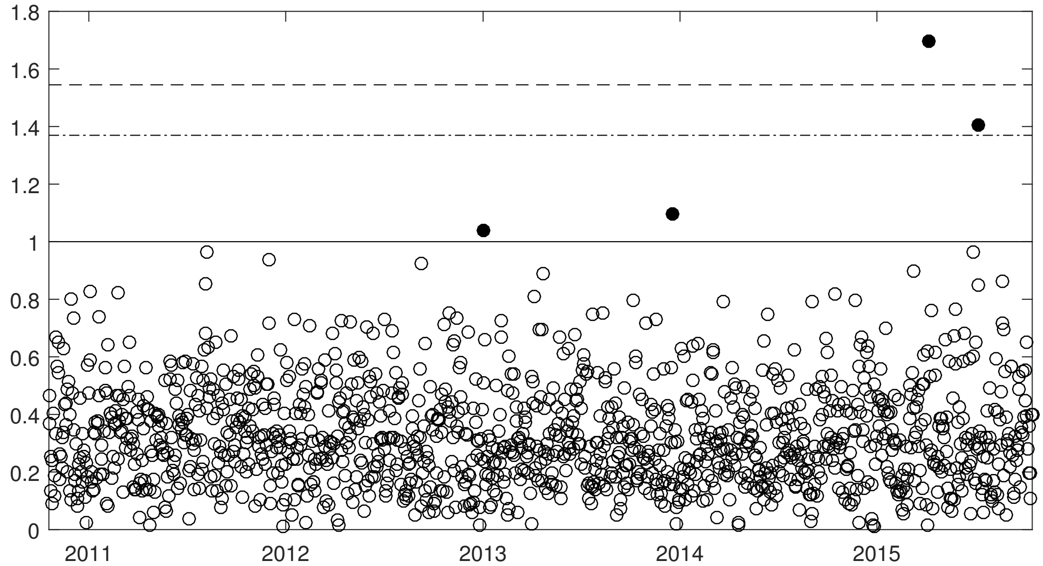


Figure 3.6: Time series of the relative distance of filtered return triples to the estimated central region with $p = 0.0027$ from Oct 15, 2010 to Oct 16, 2015 for $k = k_1 = 300$. The dashed and dash-dot lines correspond to two asymptotically conservative confidence sets of the central region respectively at the 95% and 75% confidence levels.

Another possible application of the extreme quantile region is to detect outliers. Here we define an outlier as a data point beyond the half-space depth-based central region, i.e. falls in the quantile region, with $p = 0.0027$ based on the well-known three-sigma rule in outlier detection literature. In other words, the outliers are the observations that demonstrate extremal joint behavior. While the extreme estimator of $\mathcal{Q}$ is useful in screening potential outliers, furthermore, our asymptotic approximation theory provides a possibility to detect all outliers with a certain degree of *confidence* in finite samples, by taking the statistical uncertainty into account. Figure 3.6 presents the time series of the relative distance of filtered return triples to our estimated central region with $p = 0.0027$ from October 15, 2010 to October 16, 2015 for $k = k_1 = 300$. The relative distance is calculated as the ratio of the Euclidean norm of the realized trivariate filtered return and the boundary point of our estimated central regions in the direction of the filtered return triples. We observe 4 data points above the solid horizontal line, i.e. falls in the estimated quantile region. Two out of these points are above the dash-dotted horizontal line, that is, they are outside a 75% asymptotically conservative confidence set of the true central region, but only one of them remains when comparing with the dashed horizontal line for a 95% asymptotically conservative confidence set. These two data points, in chronological order, were recorded on April 8, 2015 and July 8, 2015. We conclude that the former point is an outlier at the 95% confidence level, as well as the latter point but only at the 75% confidence level. Their outlyingness are both mainly driven by some sudden large movements in Hong Kong's stock market. The other two detected outliers (with confidence levels below 75%) are recorded on January 2 and December 19 both in 2013: the first one may attribute to the so-called January effect and the second one follows a reduction in monetary stimulus measures announced by the U.S. Federal Reserve.

3.8 Auxiliary Lemmas

Lemma 3.8.1 (Measurability)**.** *Under Assumption M, for all $\varepsilon > 0$, the functions $\mathbb{1}[\varepsilon\mathcal{Q} \subset \widehat{\mathcal{Q}}]$, $\mathbb{1}[\widehat{\mathcal{Q}} \subset \varepsilon\mathcal{Q}]$ are measurable.*

Proof. We only prove the first part; the proof of the other is completely analogous. By assumption $\widehat{\mathcal{Q}}^c$ and $\mathcal{Q}^c$ are non-empty, convex and compact. Define their *support functions* by

$$h_{\widehat{\mathcal{Q}}^c}(\mathbf{u}) = \sup\left\{\mathbf{u}^T\mathbf{x} : \mathbf{x} \in \widehat{\mathcal{Q}}^c\right\}, \quad h_{\mathcal{Q}^c}(\mathbf{u}) = \sup\left\{\mathbf{u}^T\mathbf{x} : \mathbf{x} \in \mathcal{Q}^c\right\}, \quad \mathbf{u} \in \mathbb{R}^d.$$

which are real valued and *continuous*. Furthermore, we can write

$$\widehat{\mathcal{Q}}^c = \cap_{\mathbf{u}\in\Theta}\left\{\mathbf{u}^T\mathbf{x} \le h_{\widehat{\mathcal{Q}}^c}(\mathbf{u})\right\}, \quad \mathcal{Q}^c = \cap_{\mathbf{u}\in\Theta}\left\{\mathbf{u}^T\mathbf{x} \le h_{\mathcal{Q}^c}(\mathbf{u})\right\}$$

Recall Θ_0 is a dense, countable subset of Θ. It is easy to verify that

$$\begin{aligned} \varepsilon\mathcal{Q} \subset \widehat{\mathcal{Q}} \Leftrightarrow \varepsilon\mathcal{Q}^c \supset \widehat{\mathcal{Q}}^c \Leftrightarrow h_{\widehat{\mathcal{Q}}^c}(\mathbf{u}) \le \varepsilon h_{\mathcal{Q}^c}(\mathbf{u}), \forall \mathbf{u} \in \Theta \\ \Leftrightarrow h_{\widehat{\mathcal{Q}}^c}(\mathbf{u}) \le \varepsilon h_{\mathcal{Q}^c}(\mathbf{u}), \forall \mathbf{u} \in \Theta_0, \end{aligned}$$

where the last step follows by the continuity of support functions. Now, the measurability of $\mathbb{1}[\varepsilon\mathcal{Q} \subset \widehat{\mathcal{Q}}]$ follows by the (pointwise) measurability of $h_{\widehat{\mathcal{Q}}^c}(\mathbf{u})$ on each $\mathbf{u} \in \Theta_0$, since we can also write

$$h_{\widehat{\mathcal{Q}}^c}(\mathbf{u}) = \sup_{\mathbf{v}\in\Theta} \widehat{\rho}(\mathbf{v})\mathbf{u}^T\mathbf{v} = \sup_{\mathbf{v}\in\Theta_0} \widehat{\rho}(\mathbf{v})\mathbf{u}^T\mathbf{v}, \quad \mathbf{u} \in \Theta_0,$$

where $\widehat{\rho}$ is continuous and is measurable at each point $\mathbf{v} \in \Theta$. □

Lemma 3.8.2 (Universal Separability of $\bar{\mathcal{H}}_\delta^+$)**.** *For all $\delta > 0$, there exists a countable subclass $\mathcal{D}_\delta$ of $\bar{\mathcal{H}}_\delta^+$ such that for all $B \in \bar{\mathcal{H}}_\delta^+$ there exists a sequence $(D_n)_{n\in\mathbb{N}}$ in $\mathcal{D}_\delta$ with $\mathbb{1}[\mathbf{x} \in D_n] \to \mathbb{1}[\mathbf{x} \in B]$ for all $\mathbf{x} \in \mathbb{R}^d$.*

Proof. For presentation convenience, we denote $H_{r,\mathbf{0}} = r\mathbb{C}$ and hence we can write

$$\bar{\mathcal{H}}_\delta^+ := \{H_{r,\mathbf{u}} : r > \delta, \mathbf{u} \in \Theta \cup \{\mathbf{0}\}\}.$$

The elements of $\bar{\mathcal{H}}_\delta^+$ are all indexed by the parameter $(r, \mathbf{u}) \in (\delta, \infty) \times (\Theta \cup \{\mathbf{0}\})$, those with $r \in \mathbb{Q}$ and $\mathbf{u} \in \Theta_0 \cup \{\mathbf{0}\}$ then form a countable subclass $\mathcal{D}_\delta$.

Now, let $\mathbf{x} \in \mathbb{R}^d$ and take arbitrary $H_{r,\mathbf{u}} \in \bar{\mathcal{H}}_\delta^+$. We have two possibilities.

1. When $\mathbf{u} = 0$, there exists a sequence of $r_n \downarrow r$ $(n \to \infty)$ such that

$$|\mathbb{1}[\mathbf{x} \in H_{r_n,\mathbf{0}}] - \mathbb{1}[\mathbf{x} \in H_{r,\mathbf{0}}]| = \mathbb{1}\left[r < \|x\| \le r_n\right] \to 0.$$

2. When $\mathbf{u} \in \Theta$, there exists a sequence of $(H_{r_n,\mathbf{u}_n})_{n\in\Theta}$ in $\mathcal{D}_\delta$ such that $r_n \uparrow r$ and $\mathbf{u}_n \to \mathbf{u}$. We can always take in such a way that $r_n < r$ and $\|\mathbf{u}_n - \mathbf{u}\| \le (r - r_n)^2$, $\mathbf{u}_n \in \Theta$. Then

$$\begin{aligned}
&|\mathbb{1}[\mathbf{x} \in H_{r_n,\mathbf{u}_n}] - \mathbb{1}[\mathbf{x} \in H_{r,\mathbf{u}}]| \\
=&\mathbb{1}[\mathbf{x} \in H_{r_n,\mathbf{u}_n} \cap H_{r,\mathbf{u}}^c] + \mathbb{1}[\mathbf{x} \in H_{r,\mathbf{u}} \cap H_{r_n,\mathbf{u}_n}^c] \\
\le&\mathbb{1}\left[r_n + (\mathbf{u} - \mathbf{u}_n)^T\mathbf{x} \le \mathbf{u}^T\mathbf{x} < r\right] + \mathbb{1}\left[(\mathbf{u} - \mathbf{u}_n)^T\mathbf{x} > r - r_n\right] \\
\le&\mathbb{1}\left[r_n + (\mathbf{u} - \mathbf{u}_n)^T\mathbf{x} \le \mathbf{u}^T\mathbf{x} < r\right] + \mathbb{1}\left[\|\mathbf{x}\| > (r - r_n)^{-1}\right] \\
\to&0 + 0 = 0.
\end{aligned}$$

$\square$

Lemma 3.8.3. *The Gaussian process $\mathbb{Z}$ defined in (6.14) is tight.*

Proof. Consider the intrinsic semi-metric d_2 on Θ defined by

$$d_2^2(\mathbf{u}, \mathbf{v}) := E\left[(\mathbb{Z}(\mathbf{u}) - \mathbb{Z}(\mathbf{v}))^2\right] = \frac{\psi^2}{\nu(H_{1,\mathbf{u}})\nu(H_{1,\mathbf{v}})} \cdot d_\nu(H_{1,\mathbf{u}}, H_{1,\mathbf{v}}),$$

where d_ν is defined in Lemma 3.6.6.

Recalling that $\inf_{\mathbf{u}\in\Theta} \nu(H_{1,\mathbf{u}}) = \delta_0 > 0$ and all the sample paths of $W_{\nu,1}$ are bounded and uniformly d_ν-continuous, it follows that all the sample paths of $\mathbb{Z}$ are bounded and uniformly d_2-continuous. For tightness it remains to show (Θ, d_2) is totally bounded.

When $\psi = 0$, this is trivial since $\rho_2 \equiv 0$. Let $\psi > 0$ and take arbitrary $\epsilon > 0$. Note that from the proof of Lemma 6 (when applying Theorem 3 in

Einmahl, 1997) we know that $(\mathcal{H}_1, d_\nu)$ is totally bounded with $\mathcal{H}_1 := \{H_{1,\mathbf{u}} : \mathbf{u} \in \Theta\}$, that is, we can find a finite set of $H_{1,\mathbf{u}_1}, \ldots, H_{1,\mathbf{u}_n}$ such that

$$\mathcal{H}_1 = \cup_{i=1}^{n} B_{\delta_0^2 \epsilon^2/\psi^2}(H_{1,\mathbf{u}_i}; d_\nu) \tag{3.8.1}$$

where $B_{\delta_0^2\epsilon^2/\psi^2}(H_{1,\mathbf{u}_i}; d_\nu) = \{H_{1,\mathbf{v}} : d_\nu(H_{1,\mathbf{u}_i}, H_{1,\mathbf{v}}) \le \delta_0^2\epsilon^2/\psi^2, \mathbf{v} \in \Theta\}$. Denote $\Theta_\epsilon^{(i)} = \{\mathbf{v} \in \Theta : H_{1,\mathbf{v}} \in B_{\delta_0^2\epsilon^2/\psi^2}(H_{1,\mathbf{u}_i}; d_\nu)\}$. Noting that

$$d_2^2(\mathbf{u}, \mathbf{v}) = \psi^2 \cdot \frac{d_\nu(H_{1,\mathbf{u}}, H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{u}})\nu(H_{1,\mathbf{v}})} \le \frac{\psi^2}{\delta_0^2} d_\nu(H_{1,\mathbf{u}}, H_{1,\mathbf{v}}),$$

we have

$$B_\epsilon(\mathbf{u}_i; \rho_2) := \{\mathbf{v} \in \Theta : d_2(\mathbf{u}_i, \mathbf{v}) \le \epsilon\} \supset \Theta_\epsilon^{(i)}.$$

Hence,

$$\Theta \supset \cup_{i=1}^{n} B_\epsilon(\mathbf{u}_i; \rho_2) \supset \cup_{i=1}^{n} \Theta_\epsilon^{(i)} = \Theta,$$

where the (last) equality is a consequence of (3.8.1). □

Lemma 3.8.4. *Under Assumption R, for all $\delta > 0$, as $n \to \infty$,*

$$\sup_{B \in \mathcal{A}_\delta} \left| \frac{n}{k_1} P\left(U\left(\frac{n}{k_1}\right) B\right) - \nu(B) \right| \to 0, \ \sup_{B \in \mathcal{A}_\delta} \left| \frac{n}{k} P\left(U\left(\frac{n}{k}\right) B\right) - \nu(B) \right| \to 0,$$

where $\mathcal{A}_\delta = \{B \cap B' : B, B' \in \bar{\mathcal{H}}_\delta^+\}$.

Proof. We shall only prove the first part by contradiction; the proof of the second part is completely analogous. Write for convenience $H_{r,\mathbf{0}} := r\mathbb{C}$, $r > 0$. Suppose this convergence does not hold uniformly. Then there exist some finite $r, r' > 0$ and a sequence

$$\mathbf{w}_n = (r_n' r, r_n'/r', \mathbf{u}_n, \mathbf{u}_n') \to \mathbf{w} = (1, 1, \mathbf{v}, \mathbf{v}'), \quad n \to \infty$$

such that

$$\frac{n}{k} P\left(U\left(\frac{n}{k}\right) A_n\right) - \nu(A_n) \not\to 0 \tag{3.8.2}$$

where $A_n = H_{r_n,\mathbf{u}_n} \cap H_{r_n',\mathbf{u}_n'} \in \mathcal{A}_\delta$. Write $A = H_{r,\mathbf{v}} \cap H_{r',\mathbf{v}'}$ and, without loss of generality, assume $r \le r'$. Take $\varepsilon_n := \|\mathbf{w}_n\|^{1/2} > 0$. For large n, we have $\varepsilon_n < 1$ and

$$P\left(U\left(\frac{n}{k}\right) A_n\right) \le P\left(U\left(\frac{n}{k}\right)(1-\varepsilon_n)(H_{r,\mathbf{u}_n} \cap H_{r',\mathbf{u}_n'})\right)$$

$$\begin{aligned}
\leq & P\left(U\left(\frac{n}{k}\right)(1-\varepsilon_n)^2 A\right) \\
& + P\left(U\left(\frac{n}{k}\right)(1-\varepsilon_n)(H_{r,\mathbf{u}_n} \cap (1-\varepsilon_n)H^c_{r,\mathbf{v}})\right) \\
& + P\left(U\left(\frac{n}{k}\right)(1-\varepsilon_n)(H_{r',\mathbf{u}'_n} \cap (1-\varepsilon_n)H^c_{r',\mathbf{v}'})\right).
\end{aligned}$$

Notice that, for large n, either $\mathbf{u}_n = \mathbf{v} = \mathbf{0}$ or both $\mathbf{u}_n, \mathbf{v} \in \Theta$ and therefore

$$H_{r,\mathbf{u}_n} \cap (1-\varepsilon_n)H^c_{r,\mathbf{v}} \subset \{\mathbf{x} : (\mathbf{u}_n - \mathbf{v})^T\mathbf{x} > \varepsilon_n r\} \subset \{\mathbf{x} : \varepsilon_n^2 \|\mathbf{x}\| > \varepsilon_n r\} \subset \varepsilon_n^{-1} r\mathbb{C},$$

where $\varepsilon_n^{-1} r\mathbb{C}$ reads as $\emptyset$ if $\varepsilon_n = 0$. Similarly, for large n,

$$H_{r',u'_n} \cap (1-\varepsilon_n)H^c_{r',v'} \subset \varepsilon_n^{-1} r'\mathbb{C} \subset \varepsilon_n^{-1} r\mathbb{C}$$

Hence, by Assumption 6,

$$\begin{aligned}
\frac{n}{k}P\left(U\left(\frac{n}{k}\right)A_n\right) \leq \frac{n}{k}P\left(U\left(\frac{n}{k}\right)(1-\varepsilon_n)^2 A\right) + \frac{2n}{k}P\left(U\left(\frac{n}{k}\right)(1-\varepsilon_n)\varepsilon_n^{-1} r\mathbb{C}\right) \\
\to \nu(A) + 2\cdot 0 = \nu(A).
\end{aligned}$$

Analogously, we can show that

$$\begin{aligned}
\frac{n}{k}P\left(U\left(\frac{n}{k}\right)A_n\right) \geq \frac{n}{k}P\left(U\left(\frac{n}{k}\right)(1+\varepsilon_n)^2 A\right) - \frac{2n}{k}P\left(U\left(\frac{n}{k}\right)(1+\varepsilon_n)\varepsilon_n^{-1} r\mathbb{C}\right) \\
\to \nu(A) - 2\cdot 0 = \nu(A).
\end{aligned}$$

Hence $\frac{n}{k}P\left(U\left(\frac{n}{k}\right)A_n\right) \to \nu(A)$. Similarly, even easier, we can also show that $\nu(A_n) \to \nu(A)$. Contradiction with (3.8.2). □

Lemma 3.8.5. *Assumptions R and 3.4.2 imply that for all $\varepsilon > 0$, as $t \to \infty$*

$$\sup_{\|\mathbf{x}\|\geq\varepsilon} \left| \frac{HD(t\mathbf{x}; P)}{\mathbb{P}(R > t)} - HD(\mathbf{x}; \nu) \right| = O(\alpha(t)).$$

Proof. Lemma 4 in He and Einmahl (2016) with $\delta = \delta_0^\gamma$ yields that

$$\begin{aligned}
& \sup_{\|\mathbf{x}\|\geq\varepsilon} \left| \frac{HD(t\mathbf{x}; P)}{\mathbb{P}(R > t)} - HD(\mathbf{x}; \nu) \right| \\
= & \sup_{\|\mathbf{x}\|\geq\varepsilon} \left| \inf_{\mathbf{u}^T\mathbf{x}\geq\delta_0^\gamma\varepsilon} \frac{P(tH_{\mathbf{u}^T\mathbf{x},\mathbf{u}})}{\mathbb{P}(R > t)} - \inf_{\mathbf{u}^T\mathbf{x}\geq\delta_0^\gamma\varepsilon} \nu(H_{\mathbf{u}^T\mathbf{x},\mathbf{u}}) \right|
\end{aligned}$$

$$\leq \sup_{r \geq \delta_0^\gamma \varepsilon, \mathbf{u} \in \Theta} \left| \frac{\mathbb{P}(X \in tH_{r,\mathbf{u}})}{\mathbb{P}(R > t)} - \nu(H_{r,\mathbf{u}}) \right| = O(\alpha(t)),$$

where the last equality can be easily shown analogously to Lemma 3.6.1, and we omit its proof. □

Lemma 3.8.6. *For all $\mathbf{u} \in \Theta$, $\phi(\mathbf{u})$ is non-empty, closed and $\phi(\mathbf{u}) \subset \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma\}$ with $\delta_0 := \inf_{\mathbf{u} \in \Theta} \nu(H_{1,\mathbf{u}})$.*

Proof. Let $\mathbf{u} \in \Theta$. Lemma 2.5.4 immediately implies that $\phi(\mathbf{u}) \subset \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma\}$. Now, with Lemma 2.5.3 we know

$$\nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) = (\mathbf{v}^T\mathbf{u})^{-1/\gamma}\nu(H_{1,\mathbf{v}})$$

is continuous in $\mathbf{v} \in \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma\}$. It follows that $\phi(\mathbf{u})$ is closed and, by Weierstrass's theorem, nonempty. □

Lemma 3.8.7. *Suppose Assumption R holds. The exponent measure ν satisfies Assumption 3.4.3 if $\phi(\mathbf{u})$ is a singleton for all $\mathbf{u} \in \Theta$.*

Proof. The proof is by contradiction. Suppose Assumption 3.4.3 is violated for some $\delta > 0$. Then on $\Theta \times \Theta_0$ there exists a sequence $(\mathbf{u}_n, \mathbf{w}_n) \to (\mathbf{u}, \mathbf{w}) \in \Theta \times \Theta$, such that $\mathbf{w}_n \notin \phi(\mathbf{u}_n, \delta)$, $\mathbf{w} \notin \phi(\mathbf{u}, \delta)$ and

$$\nu(H_{\mathbf{w}_n^T\mathbf{u}_n,\mathbf{w}_n}) - HD(\mathbf{u}_n; \nu) \to 0, \quad n \to \infty. \tag{3.8.3}$$

Note that

$$\begin{aligned} \left|(HD(\mathbf{u}_n;\nu))^{-\gamma} - (HD(\mathbf{u};\nu))^{-\gamma}\right| &\leq \sup_{\mathbf{v}\in\Theta} \left| \left(\nu\left(H_{\mathbf{v}^T\mathbf{u}_n,\mathbf{v}}\right)\right)^{-\gamma} - \left(\nu\left(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}\right)\right)^{-\gamma} \right| \\ &\leq \sup_{\mathbf{v}\in\Theta} \left|\mathbf{v}^T(\mathbf{u}_n - \mathbf{u})\right| \sup_{\mathbf{v}\in\Theta} (\nu(H_{1,\mathbf{v}}))^{-\gamma} \\ &\leq \|\mathbf{u}_n - \mathbf{u}\| \, \delta_0^{-\gamma} \to 0. \end{aligned}$$

It follows that $HD(\mathbf{u}_n; \nu) \to HD(\mathbf{u}; \nu)$.

Moreover, by Lemma 3 in He and Einmahl (2016) we know $\nu(H_{1,\cdot})$ is continuous on Θ and therefore

$$\nu(H_{\mathbf{w}_n^T\mathbf{u}_n,\mathbf{w}_n}) = (\mathbf{w}_n^T\mathbf{u}_n)^{-1/\gamma}\nu(H_{1,\mathbf{w}_n}) \to (\mathbf{w}^T\mathbf{u})^{-1/\gamma}\nu(H_{1,\mathbf{w}}) = \nu(H_{\mathbf{w}^T\mathbf{u},\mathbf{w}}).$$

Note that here we must have $\mathbf{w}^T\mathbf{u} \neq \mathbf{0}$, otherwise

$$\nu(H_{\mathbf{w}_n^T\mathbf{u}_n,\mathbf{w}_n}) \geq (\mathbf{w}_n^T\mathbf{u}_n)^{-1/\gamma}\delta_0 \to \infty$$

and then (3.8.3) cannot hold.

Now, we can conclude that $\nu(H_{\mathbf{w}^T\mathbf{u},\mathbf{w}}) = HD(\mathbf{u};\nu)$, that is, $\mathbf{w} \in \phi(\mathbf{u})$, contradiction with $\mathbf{w} \notin \phi(\mathbf{u},\delta)$. □

Lemma 3.8.8. *Under the conditions of Theorem 3.4.2 and the Skorohod-Dudley-Wichura construction of Lemma 3.6.7 with some $\delta \in (0,1)$*

$$\left\{\sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu)\right) : \mathbf{u}\in\Theta\right\}$$
$$\xrightarrow{\mathbb{P}} \left\{\inf_{\mathbf{v}\in\phi(\mathbf{u})} \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\left(\frac{W_{\nu,1}(H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{v}})} - W_{\nu,1}(\mathbb{C})\right) : \mathbf{u}\in\Theta\right\},$$

here '$\xrightarrow{\mathbb{P}}$' denotes convergence in probability of random elements with the supremum norm; see, e.g., Dudley (2002), page 287. Note that here the processes are only equal in distribution to the original ones.

Proof. Recall from the proof of Theorem 3.4.2 that

$$\sup_{1\geq r\geq\delta_0^\gamma,\mathbf{u}\in\Theta}\left|\sqrt{k_1}\left(\widehat{\nu}_1^*(H_{r,\mathbf{u}}) - \nu(H_{r,\mathbf{u}})\right) - \nu(H_{r,\mathbf{u}})\left(\frac{W_{\nu,1}(H_{1,\mathbf{u}})}{\nu(H_{1,\mathbf{u}})} - W_{\nu,1}(\mathbb{C})\right)\right| \xrightarrow{\mathbb{P}} 0. \tag{3.8.4}$$

Take arbitrary $\delta' > 0$. We have that

$$\begin{aligned}
&\sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu)\right)\\
=&\sqrt{k_1}\min\Big\{\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\cap\phi(\mathbf{u},\delta')} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu),\\
&\qquad\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\setminus\phi(\mathbf{u},\delta')} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu)\Big\},
\end{aligned}$$

Now with Assumption 3.4.3, uniformly on $\mathbf{u}\in\Theta$

$$\sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\setminus\phi(\mathbf{u},\delta')} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu)\right) \xrightarrow{\mathbb{P}} \infty.$$

We also have that, recalling (3.6.12) and (3.8.4), uniformly on $\mathbf{u} \in \Theta$

$$
\begin{aligned}
&\sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\cap\phi(\mathbf{u},\delta')} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu)\right)\\
&= \sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\cap\phi(\mathbf{u},\delta')} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - \inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0} \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\right)\\
&\geq \inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\cap\phi(\mathbf{u},\delta')} \sqrt{k_1}\left(\widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\right)\\
&\geq \inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\phi(\mathbf{u},\delta')} \sqrt{k_1}\left(\widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\right)\\
&\xrightarrow{\mathbb{P}} \inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\phi(\mathbf{u},\delta')} \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\left(\frac{W_{\nu,1}(H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{v}})} - W_{\nu,1}(\mathbb{C})\right).
\end{aligned}
$$

Note that Lemma 3 in He and Einmahl (2016) states that $\nu(H_{1,\cdot})$ is uniformly continuous on Θ. This implies that $\nu(H_{\mathbf{v}_1^T\mathbf{u},\mathbf{v}_1}) = (\mathbf{v}_1^T\mathbf{u})^{-1/\gamma}\nu(H_{1,\mathbf{v}_1})$ and $\nu(H_{\mathbf{v}_2^T\mathbf{u},\mathbf{v}_2}) = (\mathbf{v}_2^T\mathbf{u})^{-1/\gamma}\nu(H_{1,\mathbf{v}_2})$ can be arbitrarily close uniformly for $\mathbf{u} \in \Theta$ if $\mathbf{v}_1, \mathbf{v}_2 \in \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \varepsilon\}$ are close enough with any $\varepsilon > 0$. It follows that, for all $n \in \mathbb{N}$, there exists $\delta_n < \delta'$ such that $\delta_n \to 0$ and

$$
\phi(\mathbf{u},\delta_n) \subset \left\{\mathbf{v}\in\Theta : \left|HD(\mathbf{u};\nu) - \nu(H_{\mathbf{v}^T\mathbf{u};\mathbf{v}})\right| < \frac{1}{k_1(n)}\right\}, \quad \text{for all } \mathbf{u}\in\Theta.
$$

Hence, noting that $\phi(\mathbf{u}) \subset \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma\}$, uniformly on $\mathbf{u} \in \Theta$

$$
\begin{aligned}
&\sqrt{k_1}\left(\inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\cap\phi(\mathbf{u},\delta')} \widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - HD(\mathbf{u};\nu)\right)\\
&\leq \inf_{\mathbf{v}^T\mathbf{u}\geq\delta_0^\gamma,\mathbf{v}\in\Theta_0\cap\phi(\mathbf{u},\delta_n)} \sqrt{k_1}\left(\widehat{\nu}_1^*(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}}) - \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\right) + 1/\sqrt{k_1}\\
&\xrightarrow{\mathbb{P}} \inf_{\mathbf{v}\in\phi(\mathbf{u})} \nu(H_{\mathbf{v}^T\mathbf{u},\mathbf{v}})\left(\frac{W_{\nu,1}(H_{1,\mathbf{v}})}{\nu(H_{1,\mathbf{v}})} - W_{\nu,1}(\mathbb{C})\right),
\end{aligned}
$$

where the last step follows by (3.8.4). Since δ' is arbitrary, the claim follows since $\phi(\mathbf{u}) = \cap_{\delta'>0}\{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma, \mathbf{v} \in \phi(\mathbf{u},\delta')\}$ recalling that $\phi(\mathbf{u})$ is compact and $\phi \subset \{\mathbf{v} \in \Theta : \mathbf{v}^T\mathbf{u} \geq \delta_0^\gamma\}$ (see Section 4.2). $\square$

Chapter 4

Statistical Inference for a Relative Risk Measure

[Based on a joint work with Yanxi Hou, Liang Peng and Jiliang Sheng.]

Abstract. For monitoring systemic risk from regulators' point of view, this paper proposes a relative risk measure, which is sensitive to the market comovement. The asymptotic normality of a nonparametric estimator and its smoothed version is established. In order to effectively construct an interval estimation without complicated asymptotic variance estimation, a jackknife empirical likelihood inference procedure based on the smoothed nonparametric estimation is provided with a Wilks type of result. A simulation study and real-life data analysis show that the proposed relative risk measure is useful in monitoring systemic risk.

Key words. Copula, comovement, expected shortfall, jackknife empirical likelihood, nonparametric estimation, systemic risk

4.1 Introduction

Let X and Y denote the random losses, respectively, on an individual portfolio and some benchmark variable, say, a financial market index with joint distribution function $F(x, y)$. Consider the commonly employed expected shortfall risk measure, at level $\alpha \in (0, 1)$, defined as

$$ES_\alpha(X) = E[X|F_1(X) > 1 - \alpha] \quad \text{and} \quad ES_\alpha(Y) = E[Y|F_2(Y) > 1 - \alpha],$$

where F_1 and F_2 are the marginal distributions of X and Y given by $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$. A quick way to compare these two risk measures is to look at their ratio $ES_\alpha(X)/ES_\alpha(Y)$ (or difference). However this ratio or difference is invariant to the copula of X and Y, i.e., it is irrelevant to the market comovement. To capture the extreme dependence between X and Y, recently, Agarwal et al. (2016) proposed to multiply the above ratio by the coefficient of (upper) tail dependence (Sibuya, 1959)

$$\lambda = \lim_{t \downarrow 0} \mathbb{P}(F_1(X) > 1 - t | F_2(Y) > 1 - t),$$

which is widely studied in modeling extreme events. One may question whether we should combine this limiting tail dependence measure with the expected shortfalls calculated at a finite level α strictly larger than zero. A more natural way we propose here is to define both tail sensitivity and risk measures at the same level and this results in the following *relative risk* measure

$$\rho_\alpha = \rho_\alpha(X, Y) = \mathbb{P}(F_1(X) > 1 - \alpha | F_2(Y) > 1 - \alpha) \frac{ES_\alpha(X)}{ES_\alpha(Y)}.$$

As remarked by Agarwal et al. (2016), this relative risk measure may be viewed as an analogue of the market beta in the context of the CAPM (van Oordt and Zhou, 2016), the M-squared measure (Modigliani and Modigliani, 1997) and the Graham and Harvey's GH1 and GH2 (Graham and Harvey, 1996, 1997) measures for portfolio performance evaluation.

In order to implement the above relative risk measure ρ_α at a fixed level $\alpha \in (0,1)$, this paper first proposes a nonparametric estimation and its smoothing version and derives an asymptotic normality result. Since the asymptotic variance is quite complicated, we further investigate the possibility of employing an empirical likelihood method to construct a confidence interval since the empirical likelihood method has shown to be quite useful in interval estimation and hypothesis testing. We refer to Owen (2001) for an overview of the method. Quantifying uncertainty is important in risk management, and applications of empirical likelihood methods to risk measures have appeared in Baysal and Staum (2008); Peng et al. (2012, 2015); Wang and Peng (2016). In general, an empirical likelihood method is quite effective for linear functionals and requires linearization for a nonlinear functional by introducing some nuisance parameters. Unfortunately, it is hard to linearize the proposed relative risk measure. Therefore, we employ the smoothed jackknife empirical likelihood method to construct a confidence interval for the proposed relative risk measure as the study for copulas and tail copulas in Peng and Qi (2010) and Peng et al. (2012). Note that smoothed jackknife empirical likelihood method is a generalization of the jackknife empirical likelihood method proposed by Jing et al. (2009) for dealing with nonlinear functionals, and smoothing seems necessary for a non-smoothing nonlinear functional.

When the level α is close to zero, which is a key interest of regulators, and the sample size n is not large enough, it is useful to model α as a function of n. This is generally classified as two situations: intermediate level (i.e., $\alpha = \alpha_n \to 0$ and $\alpha_n n \to \infty$ as $n \to \infty$) and extreme level (i.e., $\alpha = \alpha_n \to 0$ and $n\alpha \to c \in [0,\infty)$ as $n \to \infty$). Such a divergent level relates to the so-called tail risk in financial econometrics, which plays an important role in risk management; see, e.g., Kelly and Jiang (2014). In general, an extreme level requires extrapolating outside the data range. Here we focus on the intermediate level and extend the above study for a fixed level to this case too. Like quantile estimation, we show that nonparametric estimation for the proposed relative risk has a different asymptotic limit for a fixed level and an intermediate level. However, the proposed smoothed jackknife empirical likelihood method gives a unified interval for ρ_α regardless of the level being fixed or intermediate.

We organize this paper as follows. Section 2 presents our nonparametric estimation procedure, jackknife empirical likelihood method, and asymptotic results. A simulation study is carried out in Section 3, and a data analysis in finance is provided in Section 4 to demonstrate the usefulness of the proposed relative measure in monitoring systemic risk. All proofs are deferred to Section 5.

4.2 Main Results

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed random vectors with distribution function $F(x, y)$ and marginals $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$. Order the X_i's as $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ and the Y_i's as $Y_{1:n} \le Y_{2:n} \le \ldots \le Y_{n:n}$. Define the survival functions $\bar{F}_i(\cdot) = 1 - F_i(\cdot)$ and quantile functions $Q_i(\cdot) = F_i^{\leftarrow}(\cdot)$ for $i = 1, 2$, where $F_i^{\leftarrow}$ denotes the (generalized) inverse function of F_i. The empirical survival functions are given by

$$\bar{F}_{n1}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(X_i > x\right), \quad \bar{F}_{n2}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(Y_i > y\right), \quad x, y \in \mathbb{R}.$$

We introduce the so-called *survival* copula function

$$C(u, v) = \mathbb{P}(\bar{F}_1(X) < u, \bar{F}_2(Y) < v), \quad u, v \in [0, 1],$$

and, under weak conditions considered below, we can rewrite

$$\rho_\alpha = \frac{1}{\alpha} C(\alpha, \alpha) \frac{ES_\alpha(X)}{ES_\alpha(Y)}$$

Substituting the right-hand-side components by their empirical counterparts yields our nonparametric estimator

$$\widetilde{\rho}_\alpha = \widetilde{\rho}_\alpha(X,Y) = \frac{1}{\alpha}\widetilde{C}(\alpha,\alpha)\frac{\widetilde{ES}_\alpha(X)}{\widetilde{ES}_\alpha(Y)},$$

where, with $\lceil\cdot\rceil$ denoting the ceiling function,

$$\begin{aligned}
\widetilde{C}(\alpha,\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left[\bar{F}_{n1}(X_i)<\alpha,\bar{F}_{n2}(Y_i)<\alpha\right],\\
\widetilde{ES}_\alpha(X) &= \frac{1}{n\alpha}\sum_{i=1}^{n}X_i\mathbb{1}\left[X_i > X_{n-\lceil n\alpha\rceil:n}\right],\\
\widetilde{ES}_\alpha(Y) &= \frac{1}{n\alpha}\sum_{i=1}^{n}Y_i\mathbb{1}\left[Y_i > Y_{n-\lceil n\alpha\rceil:n}\right].
\end{aligned}$$

Like smooth distribution (copula) estimation, we may consider a smooth version of the above nonparametric estimation. More specifically, with some density function k, its distribution function $K(x)=\int_{-\infty}^{x}k(s)ds$ and bandwidth $h=h(n)>0$, a smoothed estimator of ρ_α is given by

$$\widehat{\rho}_\alpha = \frac{1}{\alpha}\widehat{C}(\alpha,\alpha)\frac{\widehat{ES}_\alpha(X)}{\widehat{ES}_\alpha(Y)},$$

where

$$\begin{cases}
\widehat{C}(\alpha,\alpha) = \frac{1}{n}\sum_{i=1}^{n}K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2}(Y_i)/\alpha}{h}\right),\\
\widehat{ES}_\alpha(X) = \frac{1}{n\alpha}\sum_{i=1}^{n}(X_i - X_{n-\lceil n\alpha\rceil:n})K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) + X_{n-\lceil n\alpha\rceil:n},\\
\widehat{ES}_\alpha(Y) = \frac{1}{n\alpha}\sum_{i=1}^{n}(Y_i - Y_{n-\lceil n\alpha\rceil:n})K\left(\frac{1-\bar{F}_{n2}(Y_i)/\alpha}{h}\right) + Y_{n-\lceil n\alpha\rceil:n}.
\end{cases}$$

To establish the asymptotic normality of $\widetilde{\rho}_\alpha$ and $\widehat{\rho}_\alpha$ for any fixed level $\alpha\in(0,1)$, we will need the following regularity conditions.

Assumption 4.2.F (Fixed level).

(4.2.F.a) For $j=1,2$, Q_j is Lipschitz continuous in a neighborhood of $1-\alpha$ with $Q_j(1-\alpha)>0$, and F_j is strictly increasing and differentiable in a neighborhood of $Q_j(1-\alpha)$. Moreover, for some $\delta>0$, $\mathbb{E}(X_+^{2+\delta})<\infty$ and $\mathbb{E}(Y_+^{2+\delta})<\infty$, where $x_+=\max\{x,0\}$.

(4.2.F.b) C has continuous first-order derivatives $\dot{C}_1(x,\alpha) = \frac{\partial C(x,\alpha)}{\partial x}$ and $\dot{C}_2(\alpha,y) = \frac{\partial C(\alpha,y)}{\partial y}$ in a neighborhood of, respectively, $x=\alpha$ and of $y=\alpha$.

Assumption (4.2.F.a) contains standard conditions, which require underlying local continuity of the marginal distributions together with finite moments for the positive losses; see, e.g., Chen (2008). Assumption (4.2.F.b) ensures the application of the standard empirical copula process result; see, e.g., Section V in Gaenssler and Stute (1987). Below is an asymptotic normality result, where '$\xrightarrow{d}$' denotes convergence in distribution and '$\xrightarrow{\mathbb{P}}$' denotes convergence in probability.

Theorem 4.2.1 (Fixed level)**.** *For an $\alpha \in (0,1)$ satisfying $C(\alpha,\alpha) > 0$, Assumption 4.2.F implies that*

$$\sqrt{n\alpha}\left(\frac{\widetilde{\rho}_\alpha}{\rho_\alpha} - 1\right) \xrightarrow{d} N\left(0, \sigma_\alpha^2\right),$$

as $n \to \infty$, with $\sigma_\alpha^2 = Var(\Lambda_\alpha + \Theta_{\alpha,1} - \Theta_{\alpha,2})$ and the zero-mean Gaussian random variables

$$\Lambda_\alpha = \frac{\sqrt{\alpha}}{C(\alpha,\alpha)}\left\{B_C(\alpha,\alpha) - \dot{C}_1(\alpha,\alpha)B_C(\alpha,1) - \dot{C}_2(\alpha,\alpha)B_C(1,\alpha)\right\}$$

$$\Theta_{\alpha,1} = -\frac{\frac{1}{\sqrt{\alpha}}\int_0^1 B_C(\alpha x,1)dQ_1(1-\alpha x)}{ES_\alpha(X)},$$

$$\Theta_{\alpha,2} = -\frac{\frac{1}{\sqrt{\alpha}}\int_0^1 B_C(1,\alpha y)dQ_2(1-\alpha y)}{ES_\alpha(Y)}.$$

Here, B_C is a C-Brownian bridge, i.e., a zero-mean Gaussian process with covariance function

$$\mathbb{E}(B_C(u_1,v_1)B_C(u_2,v_2)) = C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1,v_1)C(u_2,v_2), \quad (u_1,v_1),(u_2,v_2) \in [0,1]^2.$$

Furthermore, if k is a symmetric density with support [-1,1] and bounded first derivative and the bandwidth $h = h(n) > 0$ satisfies

$$nh^2 \to \infty \quad \text{and} \quad nh^4 \to 0,$$

then we have that, as $n \to \infty$,

$$\sqrt{n\alpha}\left(\frac{\widehat{\rho}_\alpha - \widetilde{\rho}_\alpha}{\rho_\alpha}\right) \xrightarrow{\mathbb{P}} 0.$$

Theorem 4.2.1 states that, under weak regularity conditions, both the non-smoothed estimator $\widetilde{\rho}_\alpha$ and smoothed estimator $\widehat{\rho}_\alpha$ are asymptotically normal with the same limiting distribution.

When α is close to zero (but not extremely), as discussed in Introduction, it is often useful to model α as an intermediate sequence of n in such a way that $\alpha = \alpha_n \to 0$ and $n\alpha_n \to \infty$, as $n \to \infty$. For the study of an intermediate level α, in the context of extreme value theory, one needs some conditions on the tail behavior of the underlying variables as follows.

Assumption 4.2.I (Intermediate level).

(4.2.I.a) For some $\gamma_j \in (0, 1/2)$, $\beta_j \leq 0$ and function A_j with a constant sign near infinity,

$$\lim_{t\to\infty} \frac{1}{A_j(1/\bar{F}_j(t))}\left(\frac{\bar{F}_j(tx)}{\bar{F}_j(t)} - x^{-1/\gamma_j}\right) = x^{-1/\gamma_j}\frac{x^{\beta_j/\gamma_j} - 1}{\gamma_j\beta_j}, \quad x > 0,$$

for all $j = 1, 2$.

(4.2.I.b) There exists a function $R : (0,\infty)^2 \to [0,\infty)$ such that

$$\lim_{t\to\infty} tC(t^{-1}x, t^{-1}y) = R(x, y), \quad (x, y) \in (0,\infty)^2, \tag{4.2.1}$$

and it has continuous first-order derivatives $\dot{R}_1(x, y) = \frac{\partial R(x,y)}{\partial x}$ and $\dot{R}_2(x, y) = \frac{\partial R(x,y)}{\partial y}$ on a neighborhood of $(1, 1)$.

(4.2.I.c) The function C has first-order derivatives $\dot{C}_1(x, y) = \frac{\partial C(x,y)}{\partial x}$ and $\dot{C}_2(x, y) = \frac{\partial C(x,y)}{\partial y}$ on $(0, \delta)^2$ for some $\delta > 0$, and, as $t \to \infty$,

$$\sup_{x,y\in(1-\delta,1+\delta)} \left|\dot{C}_1(t^{-1}x, t^{-1}y) - \dot{R}_1(x, y)\right| \to 0,$$
$$\sup_{x,y\in(1-\delta,1+\delta)} \left|\dot{C}_2(t^{-1}x, t^{-1}y) - \dot{R}_2(x, y)\right| \to 0.$$

Assumption (4.2.I.a) is a standard second order condition in univariate extreme value theory; see, e.g. Section 2.3 in de Haan and Ferreira (2006). The condition $\gamma_1, \gamma_2 < \frac{1}{2}$ implies that there exists some $\delta_1 > 0$ such that $EX_+^{2+\delta_1} < \infty$ and $EY_+^{2+\delta_1} < \infty$. Assumptions (4.2.I.b) and (4.2.I.c) can be viewed as tail analogues of Assumption (4.2.F.b) for applying the theory of tail copula process; see, e.g., Cai et al. (2015), Einmahl et al. (2006) and Theorem 7.2.2 in de Haan and Ferreira (2006). The R-function defined therein fully characterizes the so-called *stable tail dependence function* l in such a way that

$$l(x, y) = x + y - R(x, y), \quad x, y \geq 0;$$

see, e.g., Drees and Huang (1998) and Section 8.2 in Beirlant et al. (2006).

Theorem 4.2.2 (Intermediate level)**.** *Let $\alpha = \alpha_n$ be an intermediate sequence, that is, $\alpha_n \to 0$ and $n\alpha_n \to \infty$, as $n \to \infty$. Given $R(1,1) > 0$, Assumption 4.2.I implies that, as $n \to \infty$,*

$$\sqrt{n\alpha}\left(\frac{\widetilde{\rho}_\alpha}{\rho_\alpha} - 1\right) \xrightarrow{d} N\left(0, \sigma_0^2\right),$$

with $\sigma_0^2 = Var(\Lambda_0 + \Theta_{0,1} - \Theta_{0,2})$ and the zero-mean Gaussian random variables

$$\Lambda_0 = \frac{W_R(1,1) - \dot{R}_1(1,1)W_R(1,\infty) - \dot{R}_2(1,1)W_R(\infty,1)}{R(1,1)},$$

$$\Theta_{0,1} = (\gamma_1 - 1)\int_0^1 W_R(x,\infty)dx^{-\gamma_1}, \quad \Theta_{0,2} = (\gamma_2 - 1)\int_0^1 W_R(\infty,y)dy^{-\gamma_2},$$

Here, W_R is a R-Brownian motion, i.e. a zero-mean Gaussian process with covariance function

$$\mathbb{E}(W_R(u_1,v_1)W_R(u_2,v_2)) = R(u_1 \wedge u_2, v_1 \wedge v_2) \quad \textit{for} \quad (u_1,v_1),(u_2,v_2) \in (0,\infty]^2 \backslash \{\infty,\infty\}.$$

Furthermore, if k is a symmetric density with support [-1,1] and bounded first derivative and the bandwidth $h = h(n) > 0$ satisfies

$$n\alpha h^2 \to \infty, \quad n\alpha h^4 \to 0, \quad \textit{and} \quad n\alpha h^2 A_i^2(1/\alpha) = O(1) \quad \textit{for} \quad i = 1, 2,$$

then we have that, as $n \to \infty$,

$$\sqrt{n\alpha}\left(\frac{\widehat{\rho}_\alpha - \widetilde{\rho}_\alpha}{\rho_\alpha}\right) \xrightarrow{\mathbb{P}} 0.$$

Theorem 4.2.2 is a tail analogue of Theorem 4.2.1, despite slightly stronger conditions are imposed to eliminate the asymptotic bias due to the extreme-value approximations.

Based on these two asymptotic normality results, one can construct a confidence interval of ρ_α based on either $\widetilde{\rho}_\alpha$ or $\widehat{\rho}_\alpha$. Estimating the asymptotic variance of $\widetilde{\rho}_\alpha$ or $\widehat{\rho}_\alpha$ requires some (empirical) approximation of the copula function C or the function R, say, $\widehat{C}$ and $\widehat{R}$ respectively. A usual approach requires simulating the Gaussian process $B_{\widehat{C}}$ or $W_{\widehat{R}}$ with some empirical approximations of the limiting covariance functions. It is also necessary to estimate the first-order partial derivatives of the (tail) copula function and even, when α_n is an intermediate sequence, the tail indices of the marginal distributions. This approach is often quite computationally intensive, and its finite-sample performance can be quite poor by aggregating all the estimation uncertainties discussed above.

Instead, we investigate the possibility of employing the empirical likelihood method. Although this method proposed by Owen (1988) and Owen (1990) has proved to be quite effective in interval estimation and hypothesis testing, it has a serious problem in handling a nonlinear statistic. For example, it can lead to computational difficulties by solving a number (dependent on n) of simultaneous equations. Recently Jing et al. (2009) proposed a so-called jackknife empirical likelihood method for dealing with nonlinear statistics such as U-statistics. Like inference for receiver operating characteristic (ROC) curves, copulas and tail copulas in Gong et al. (2010), Peng et al. (2012) and Peng and Qi (2010), a smoothed version is needed for the proposed relative risk measure.

Hence we shall establish our jackknife empirical likelihood inference method for ρ_α based on the smoothed nonparametric estimation. To apply the smoothed jackknife empirical likelihood method, we first need to construct a jackknife pseudo sample of ρ_α given by

$$\widehat{V}_{\rho,i} = n\widehat{\rho}_\alpha - (n-1)\widehat{\rho}_{\alpha,i}, \quad i = 1, \ldots, n,$$

where

$$\widehat{\rho}_{\alpha,i} = \frac{1}{\alpha}\widehat{C}_i(\alpha,\alpha)\frac{\widehat{ES}_{\alpha,i}(X)}{\widehat{ES}_{\alpha,i}(Y)}$$

with

$$\begin{cases} \widehat{C}_i(\alpha,\alpha) = \frac{1}{n-1}\sum_{j\neq i} K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2,i}(Y_j)/\alpha}{h}\right), \\ \widehat{ES}_{\alpha,i}(X) = \frac{1}{(n-1)\alpha}\sum_{j\neq i}(X_j - X_{n-\lceil n\alpha\rceil:n})K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) + X_{n-\lceil n\alpha\rceil:n}, \\ \widehat{ES}_{\alpha,i}(Y) = \frac{1}{(n-1)\alpha}\sum_{j\neq i}(Y_j - Y_{n-\lceil n\alpha\rceil:n})K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) + Y_{n-\lceil n\alpha\rceil:n}, \end{cases}$$

and

$$\bar{F}_{n1,i}(x) = \frac{1}{n-1}\sum_{j\neq i}\mathbb{1}\left[X_j > x\right], \quad \bar{F}_{n2,i}(y) = \frac{1}{n-1}\sum_{j\neq i}\mathbb{1}\left[Y_j > y\right], \quad x, y \in \mathbb{R}.$$

Based on this pseudo sample, the jackknife empirical likelihood ratio function for $\theta = \rho_\alpha$ can be defined by

$$\widehat{\mathcal{R}}(\theta) = \sup\left\{\Pi_{i=1}^n np_i : p_1 > 0, \ldots, p_n > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\widehat{V}_{\rho,i} = \theta\right\}.$$

Applying the Lagrange multiplier method yields

$$p_i = \frac{1}{n}\frac{1}{1+\lambda(\widehat{V}_{\rho,i}-\theta)}, \tag{4.2.2}$$

where $\lambda = \lambda(\theta)$ solves the equation

$$\frac{1}{n}\sum_{i=1}^n \frac{\widehat{V}_{\rho,i}-\theta}{1+\lambda(\widehat{V}_{\rho,i}-\theta)} = 0. \tag{4.2.3}$$

It follows that the log empirical likelihood ratio is

$$-2\log\widehat{\mathcal{R}}(\theta) = 2\sum_{i=1}^n \log\left\{1+\lambda(\widehat{V}_{\rho,i}-\theta)\right\}.$$

To unify our JEL procedure for fixed and intermediate level α, we need one more assumption.

Assumption 4.2.C. For some $\tau > \max\{\gamma_1, \gamma_2\}$ such that

$$\lim_{t \downarrow 0} \sup_{0<x,y \le 1} \frac{|t^{-1}C(tx,ty) - R(x,y)|}{(x \wedge y)^{\tau}} = 0, \tag{4.2.4}$$

where $x \wedge y := \min\{x, y\}$.

This is very similar to the condition (a) in Cai et al. (2015) but we allow an arbitrary rate of convergence here. We can show that (4.2.4) is satisfied with $\tau = 1$ if

$$\lim_{t \downarrow 0} \sup_{x \ge 1} \left| t^{-1}C(tx,t) - R(x,1) \right| = 0 \text{ and } \lim_{t \downarrow 0} \sup_{y \ge 1} \left| t^{-1}C(t,ty) - R(1,y) \right| = 0,$$

which is weaker than the usual second-order condition (7.2.8) in de Haan and Ferreira (2006).

Below is a Wilks type result for our JEL approach.

Theorem 4.2.3. *Either if the conditions of Theorem 4.2.1 hold, or if the conditions of Theorem 4.2.2 in conjunction with Assumption 4.2.C hold, we have, as $n \to \infty$,*

$$-2 \log \widehat{\mathcal{R}}(\rho_\alpha) \xrightarrow{d} \chi_1^2.$$

Based on Theorem 4.2.3, an asymptotic confidence interval with level ψ for ρ_α is given by

$$I_\psi = \{\theta \in \mathbb{R} : -2 \log \widehat{\mathcal{R}}(\theta) \le \chi^2_{1,\psi}\}$$

where $\chi^2_{1,\psi}$ is the ψ-th quantile of the chi-squared distribution with 1 degree of freedom. This interval has the asymptotically correct coverage probability regardless of the level α being fixed or intermediate. In other words, for certain sample size n and small level α, both asymptotic embeddings lead to the same approximation. This interval can be efficiently determined using a standard search algorithm; for more details we refer to Section 2.9 in Owen (2001).

4.3 Simulation Study

In this section, a simulation study is carried out to evaluate the finite-sample behavior of the proposed jackknife empirical likelihood method for our proposed relative risk measure ρ_α. The survival copula in our simulation study is a so-called t-copula with multiple parameters of degrees of freedom which is a generalization of the grouped t-copula; see Luo and Shevchenko (2010) for details. The distribution of a two-dimensional t-copula with multiple parameters of degrees of freedom is

$$C^{\Sigma}_{\nu_1,\nu_2}(u_1,u_2)=\int_0^1 \Phi_\Sigma(z_1(u_1,s),z_2(u_2,s))ds, \quad u_1,u_2\in[0,1],$$

- Φ_Σ is the distribution function of a bivariate normal random vector with zero means, unit variances and positive correlation ρ;
- $z_i(u_i,s)=t^{-1}_{\nu_i}(u_i)/\omega_i(s), \omega_i(s)=\sqrt{\nu_i/\chi^{-1}_{\nu_i}(s)}$, $i=1,2$;
- t_{ν_i} and $t^{-1}_{\nu_i}$ denote the distribution function and quantile of a student-t random variable with ν_i degrees of freedom respectively, $i=1,2$;
- χ_{ν_i} and $\chi^{-1}_{\nu_i}$ denote the distribution function and quantile of a chi-squared random variable with ν_i degrees of freedom respectively, $i=1,2$.

We draw 1000 random samples of size $n=500$ and 1000 from a bivariate distribution with a t-copula with two parameters of degrees of freedom $\nu=(\nu_1,\nu_2)\in\{(3,3),(3,5),(5,3),(5,5)\}$ and two marginal t distributions with degrees of freedom ν_1 and ν_2 respectively. We consider two cut-off levels $\alpha=0.05,0.1$ and two confidence levels $\psi=0.9,0.95$. In all cases, we set $\rho=0.2$.

The empirical coverage probability of the jackknife empirical likelihood-based confidence interval is compared to that of the bootstrap confidence interval. The bootstrap confidence interval is obtained by using 1000 bootstrap samples of size n from each sample $X_1,\ldots,X_n$. Specifically, for each

Table 4.1: Empirical coverage probabilities for the jackknife empirical likelihood-based confidence interval $I_\psi(h)$ and the bootstrap confidence interval I_ψ^* of ρ_α with cutoff level $\boldsymbol{\alpha = 0.05}$, sample size $n = 500, 1000$ and confidence levels $\psi = 0.90, 0.95$. Bandwidths are chosen as $h_1 = 0.5(n\alpha)^{-\frac{1}{3}}$, $h_2 = (n\alpha)^{-\frac{1}{3}}$, $h_3 = 1.5(n\alpha)^{-\frac{1}{3}}$, $h_4 = 2(n\alpha)^{-\frac{1}{3}}$, $h_5 = 2.5(n\alpha)^{-\frac{1}{3}}$, $h_6 = 3(n\alpha)^{-\frac{1}{3}}$.

	$n = 500$				$n = 1000$			
(ν_1, ν_2)	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$
$I_{0.95}^*$	0.880	0.873	0.882	0.860	0.920	0.922	0.914	0.919
$I_{0.95}(h_1)$	0.938	0.954	0.939	0.945	0.934	0.934	0.937	0.943
$I_{0.95}(h_2)$	0.947	0.959	0.953	0.960	0.947	0.945	0.951	0.939
$I_{0.95}(h_3)$	0.945	0.954	0.956	0.958	0.945	0.940	0.955	0.946
$I_{0.95}(h_4)$	0.940	0.950	0.952	0.955	0.932	0.932	0.935	0.939
$I_{0.95}(h_5)$	0.941	0.944	0.949	0.955	0.927	0.934	0.933	0.935
$I_{0.95}(h_6)$	0.925	0.938	0.947	0.950	0.923	0.924	0.927	0.932
$I_{0.90}^*$	0.834	0.824	0.838	0.831	0.868	0.870	0.874	0.867
$I_{0.90}(h_1)$	0.890	0.903	0.886	0.889	0.883	0.897	0.892	0.890
$I_{0.90}(h_2)$	0.900	0.914	0.911	0.915	0.899	0.892	0.890	0.900
$I_{0.90}(h_3)$	0.896	0.903	0.903	0.910	0.902	0.893	0.899	0.898
$I_{0.90}(h_4)$	0.885	0.899	0.905	0.900	0.884	0.881	0.885	0.889
$I_{0.90}(h_5)$	0.880	0.892	0.901	0.903	0.879	0.876	0.887	0.890
$I_{0.90}(h_6)$	0.865	0.882	0.885	0.896	0.875	0.873	0.876	0.889

bootstrap sample, we calculate the empirical estimate of ρ_α, which results in 1000 bootstrapped empirical estimates of ρ_α, denoted as $\widetilde{\rho}_\alpha^{*1}, \ldots, \widetilde{\rho}_\alpha^{*1000}$, and therefore 1000 bootstrap differences $\delta^{*i} = \widetilde{\rho}_\alpha^{*i} - \rho_\alpha$, $i = 1, \ldots, 1000$. Ordering these bootstrap differences by $\delta^{*[1]} \leq \ldots \delta^{*[1000]}$, the bootstrap confidence interval at level ψ is then calculated as

$$I_\psi^* = [\widetilde{\rho}_\alpha - \delta^{*[n2]},\ \widetilde{\rho}_\alpha - \delta^{*[n1]}],$$

where n_1 and n_2 denotes the integer part of $500(1-\psi)$ and $500(1+\psi)$, respectively. Motivated by the optimal bandwidth choice in smoothing distribution function estimation, we choose $h = d(n\alpha)^{-1/3}$ for various $d = 0.5, 1, 1.5, 2, 2.5, 3$.

We report the empirical coverage probabilities in Tables 4.1 and 4.2, which show that the proposed jackknife empirical likelihood method performs better than the bootstrap method in terms of coverage accuracy, and the results are quite stable with respect to the different choices of bandwidth h especially with $d = 1, 1.5, 2$. For $\alpha = 0.1$, it clearly shows that a larger size improves the accuracy.

4.4 Real-life Data Analysis

In this section, we study our relative risk measure in a real-life data set, which contains daily stock losses on 18 largest U.S. Banks [1] and Standard & Poor's 500 index (benchmark) between February 1st, 2002 and March 31st, 2011 from the Center for Research in Security Prices (CRSP), and weekly levels of the Adjusted National Financial Conditions Index (ANFCI) between September 1, 2006 and March 25, 2011 from the Federal Reserve Bank of Chicago. Positive values of the ANFCI indicate financial conditions (relative to the contemporaneous economic conditions) are tighter than average, while the negative values indicate that financial conditions (relative to the contemporaneous economic conditions) are looser than average. We document some empirical evidence of a nonlinear relation of our bank-specific relative risk measures to the one quarter future ANFCI in our sample period. Particularly, in our sample period, we observe a *minimal* U.S. financial *instability* on average measured by ANFCI when the banks maintain their one-quarter

[1] This list is from the 2009 Supervisory Capital Assessment Program (also known as 2009 bank stress tests). We exclude GMAC (now known as Ally Financial) because it only had preferred stock trading over our sample period.

Table 4.2: Empirical coverage probabilities for the jackknife empirical likelihood-based confidence interval $I_\psi(h)$ and the bootstrap confidence interval I^*_ψ of ρ_α with cutoff levels $\boldsymbol{\alpha = 0.1}$, sample size $n = 500, 1000$ and confidence levels $\psi = 0.90, 0.95$. Bandwidths are chosen as $h_1 = 0.5(n\alpha)^{-\frac{1}{3}}$, $h_2 = (n\alpha)^{-\frac{1}{3}}$, $h_3 = 1.5(n\alpha)^{-\frac{1}{3}}$, $h_4 = 2(n\alpha)^{-\frac{1}{3}}$, $h_5 = 2.5(n\alpha)^{-\frac{1}{3}}$, $h_6 = 3(n\alpha)^{-\frac{1}{3}}$.

	$n = 500$				$n = 1000$			
(ν_1, ν_2)	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$	$(3,3)$	$(3,5)$	$(5,3)$	$(5,5)$
$I^*_{0.95}$	0.920	0.922	0.914	0.919	0.935	0.939	0.937	0.935
$I_{0.95}(h_1)$	0.930	0.940	0.938	0.934	0.950	0.943	0.948	0.953
$I_{0.95}(h_2)$	0.938	0.946	0.947	0.950	0.948	0.946	0.954	0.959
$I_{0.95}(h_3)$	0.942	0.949	0.944	0.947	0.946	0.947	0.950	0.956
$I_{0.95}(h_4)$	0.932	0.932	0.935	0.939	0.951	0.952	0.943	0.946
$I_{0.95}(h_5)$	0.927	0.934	0.933	0.935	0.948	0.947	0.937	0.946
$I_{0.95}(h_6)$	0.923	0.924	0.927	0.932	0.943	0.947	0.939	0.940
$I^*_{0.90}$	0.868	0.870	0.874	0.867	0.874	0.887	0.874	0.890
$I_{0.90}(h_1)$	0.880	0.885	0.877	0.882	0.911	0.890	0.906	0.905
$I_{0.90}(h_2)$	0.884	0.891	0.884	0.891	0.909	0.906	0.901	0.904
$I_{0.90}(h_3)$	0.879	0.894	0.888	0.891	0.910	0.903	0.897	0.907
$I_{0.90}(h_4)$	0.884	0.881	0.885	0.889	0.897	0.903	0.893	0.896
$I_{0.90}(h_5)$	0.879	0.876	0.887	0.890	0.887	0.899	0.891	0.899
$I_{0.90}(h_6)$	0.875	0.873	0.876	0.889	0.880	0.892	0.886	0.893

lagged relative risk at nearly *unit* level. Therefore our inference method may serve as a practical statistical tool to monitor whether the institutional relative risks deviate significantly from their 'optimal' levels (if there are indeed any) at a certain confidence level.

As usual, our collected stock returns exhibit so-called *volatility clustering* behavior widely documented in the empirical finance literature: the univariate squared stock returns are moderately auto-correlated. Hence, we shall work on a filtered version of the univariate losses. We use a rolling window size of 1155 days (half of our daily sample size), that is, for each day we look at the last 1155 days (roughly 5 years) daily data and calibrate a generalized autoregressive conditional heteroskedasticity GARCH(1,1) model proposed in Bollerslev (1986) for each time series. Specifically, we consider the model

$$X_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = a_0 + a_1 X_{t-1}^2 + b_1 \sigma_{t-1}^2,$$
$$Y_t = \widetilde{\sigma}_t \widetilde{\epsilon}_t, \quad \widetilde{\sigma}_t^2 = \widetilde{a}_0 + \widetilde{a}_1 Y_{t-1}^2 + \widetilde{b}_1 \widetilde{\sigma}_{t-1}^2,$$

where X_t is the stock loss on an individual institution, Y_t is the benchmark (loss on S&P 500 in our study), $a_0, a_1, b_1, \widetilde{a}_0, \widetilde{a}_1, \widetilde{b}_1$ are nonnegative parameters, and $(\epsilon_t, \widetilde{\epsilon}_t)'s$ are i.i.d. innovations with zero means and unit variances. Given the past observations $(X_1, Y_1), \cdots, (X_t, Y_t)$, we are interested in the one-period ahead prediction of the relative risk measure at $t+1$, that is,

$$\rho_{\alpha,t+1|t} = \rho_\alpha(X_{t+1}, Y_{t+1} | X_1, Y_1, \ldots, X_t, Y_t) = \frac{\sigma_{t+1}}{\widetilde{\sigma}_{t+1}} \rho_\alpha^\epsilon, \tag{4.4.1}$$

where $\rho_\alpha^\epsilon = \rho_\alpha(\epsilon_t, \widetilde{\epsilon}_t)$ is the (unconditional) relative risk measure of ϵ_t against $\widetilde{\epsilon}_t$. The parameters in the GARCH models are estimated by maximizing the quasi-likelihood corresponding to normally distributed innovations, and therefore, a sample version of the innovations are obtained. For simplicity, we shall treat all these sample innovations as true values in our analysis, which means that we ignore the variability of the estimates in fitting GARCH models. In the future, we shall extend the proposed jackknife empirical likelihood method for independent data in Section 2 to take into account the estimation uncertainty of fitting GARCH models. For each day, we also calculate an aggregate predicted relative risk measure as the cross-sectional average of the individual values.

Our daily prediction of relative risk starts from the second half of our sample period, that is, from September 1st, 2006 to April 1st, 2011. The aggregate predicted relative risk ranges from 0.56 to 2.39 with a sample median value 0.97 in our sample period. Figures 4.1 presents the time series of our predicted $\rho_{\alpha,t+1|t}$ with $h = 2(n\alpha)^{-1/3}$ for the largest three banks[2]: J.P. Morgan, Bank of America and Wells Fargo. The intervals, at 95% level, are constructed by firstly the smoothed jackknife empirical likelihood method on the estimated innovations and then a simple affine transformation using the relation (4.4.1). The horizontal lines refer to the *unit* relative risk level. As we see, the patterns of the proposed relative risk for these three banks are distinct, which may provide useful information in monitoring systemic risk.

In the following we shall study on the relation of our predicted relative risk measure to the ANFCI. Since ANFCI is only reported *weekly* (every Friday), we convert our (non-smoothed) predicted relative risk measure into weakly basis by averaging the daily predictions in the same week. Starting from November 24, 2006, we first run a preliminary linear regression using the one-quarter (12 weeks) lagged ANFCI level, nominal and squared weekly bank-specific relative risk at individual or aggregate level as explanatory variables and the current ANFCI level as the dependent variable. Specifically, we run the following regression with t being the week index:

$$ANFCI_t = \omega + \phi_0 ANFCI_{t-12} + \phi_1 \cdot RelativeRisk_{t-12} + \phi_2 \cdot RelativeRisk^2_{t-12} + \varepsilon_t,$$

where $RelativeRisk$ is the aggregate or individual relative risk measure; in the latter case we only report the results, like in Figure 4.1, for J.P. Morgan, Bank of America and Wells Fargo.

The second to fifth columns in Panel A of Table 4.3 present the OLS estimates and t-statistics robust to heteroskedasticity and autocorrelation based on Newey and West (1987) and Newey and West (1994). We observe significantly negative ϕ_1's and positive ϕ_2's in general and we cannot reject neither the null hypothesis $\phi_1 + 2\phi_2 = 0$ nor $\omega + \phi_1 + \phi_2 = 0$ at 5% level by Wald test (see the last two columns in Panel A for the p-values). Motivated by these two relationship, if one replaces $\phi_1 = -2\phi_2$ and $\omega = \phi_2$, the model can be written in terms of $(RelativeRisk - 1)^2$ with a zero intercept. This

[2]The ranking is based on total assets as of March 31, 2016.

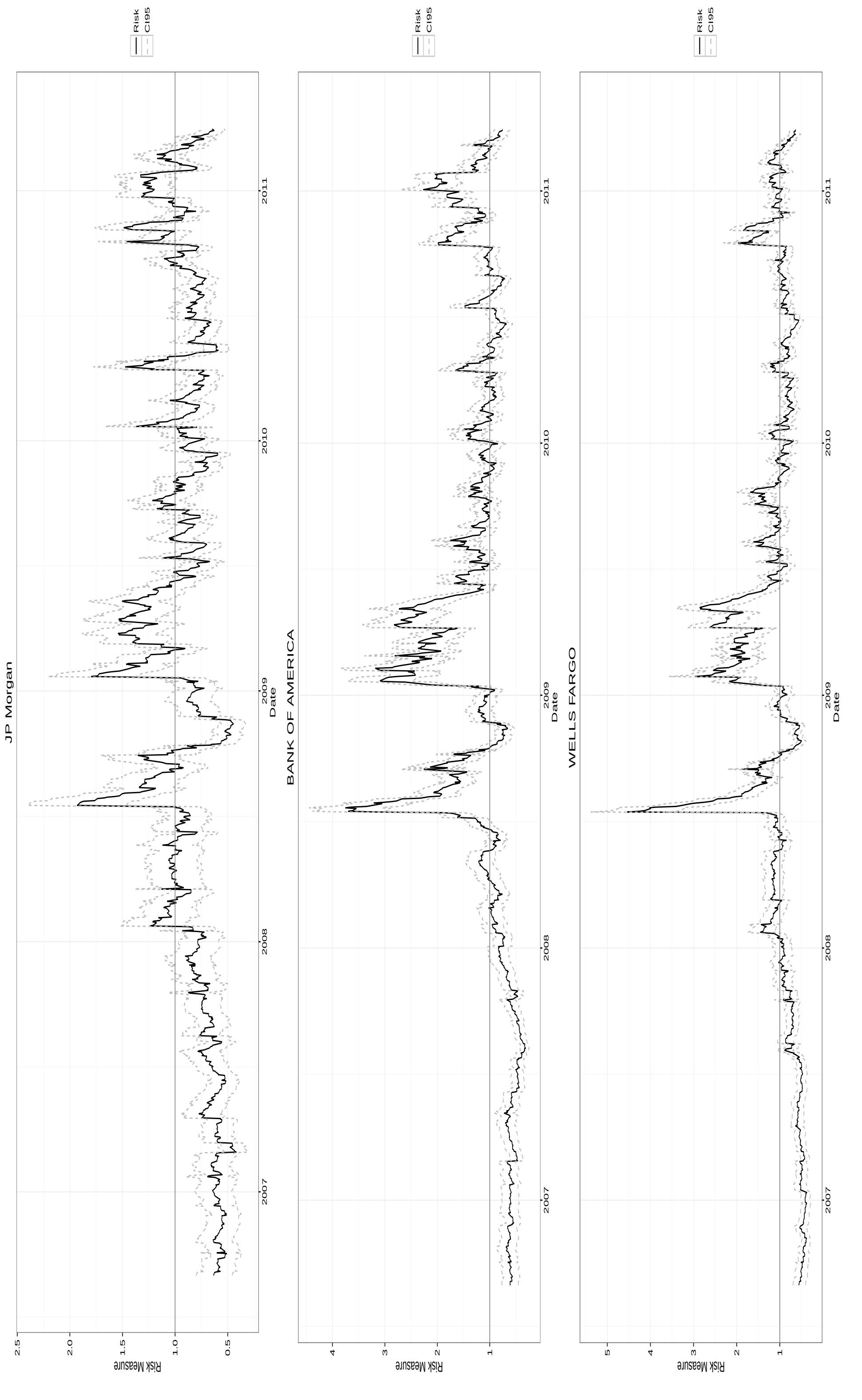


Figure 4.1: Time Series from September 1st, 2006 to April 1st, 2011 of the one-day ahead predicted values and 95% intervals of the relative tail risk with cut-off level $\alpha = 0.05$ of the daily stock losses on J.P. Morgan, Bank of America and Wells Fargo against the daily loss on Standard & Poor's 500 index. Every prediction is calculated over a 1155-days rolling window based on the smoothed estimator with bandwidth $h = 2(n\alpha)^{-1/3}$.

Table 4.3: The OLS estimates, t-statistics (in parenthesis) and Wald test p-values. The t-statistics are calculated based on the Newey-West standard errors from Newey and West (1987) with the automatic lag selection algorithm from Newey and West (1994). The null hypotheses in Wald tests are (a) $\phi_1 + 2\phi_2 = 0$ (b) $\omega + \phi_1 + \phi_2 = 0$.

Panel A: $ANFCI_t = \omega + \phi_0 ANFCI_{t-12} + \phi_1 \cdot RelativeRisk_{t-12} + \phi_2 \cdot RelativeRisk^2_{t-12} + \varepsilon_t$

	ω	ϕ_0	ϕ_1	ϕ_2	(a)	(b)
Aggregate	3.25** (2.81)	0.35** (3.83)	-6.00** (-3.03)	2.75** (3.50)	0.48	0.95
J.P. Morgan	2.88** (3.56)	0.38** (3.87)	-6.68** (-4.25)	3.86** (5.56)	0.07	0.80
Bank of America	1.54** (2.51)	0.36** (4.53)	-2.52** (-2.62)	0.99** (3.97)	0.28	0.95
Wells Fargo	0.81 (1.42)	0.38** (3.34)	-1.42 (-1.55)	0.69** (3.20)	0.94	0.72

Panel B: $ANFCI_t = \widetilde{\omega} + \widetilde{\phi}_0 ANFCI_{t-12} + \widetilde{\phi}_1 \cdot (RelativeRisk_{t-12} - 1)^2 + \widetilde{\varepsilon}_t$

	$\widetilde{\omega}$	$\widetilde{\phi}_0$	$\widetilde{\phi}_1$
Aggregate	0.04 (0.33)	0.34** (4.62)	2.20** (2.88)
J.P. Morgan	-0.07 (-0.35)	0.41** (3.84)	3.82** (3.39)
Bank of America	0.05 (0.32)	0.35** (4.11)	0.69** (4.03)
Wells Fargo	0.08 (0.46)	0.38** (4.56)	0.67** (8.19)

*significance at 5% level, **significance at 1% level.

model assumption coincides with the result of a second linear regression we run as follows, which is reported in Panel B of Table 4.3:

$$ANFCI_t = \widetilde{\omega} + \widetilde{\phi}_0 ANFCI_{t-12} + \widetilde{\phi}_1 \cdot (RelativeRisk_{t-12} - 1)^2 + \widetilde{\varepsilon}_t.$$

We observe insignificant $\widetilde{\omega}$'s but significantly positive $\widetilde{\phi}_1$'s in all cases. This suggest that ANFCI was *minimal* on average when the one-quarter lagged aggregate relative risk or individual relative risk of the largest three banks was at unit level in our sample period. The intervals from Figure 4.1 therefore can be used to monitor whether the next-period institutional relative risk may deviate significantly from their 'optimal' levels, which may be taken to be 1 here, at the 95% confidence level.

4.5 Proofs

This section starts from the asymptotics of $\widehat{\rho}_\alpha$ and $\widetilde{\rho}_\alpha$, i.e. Theorems 4.2.1 and 4.2.2. Our proofs will be based on some (well-known) asymptotic results of the weighted empirical copula process (e.g., Appendix G in Genest and Segers (2009)) and their tail analogues (see, e.g., Lemma 1 in Cai et al. (2015)). Define the *pseudo* estimator of the survival copula function by

$$\check{C}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left[\bar{F}_1(X_i) < x, \bar{F}_2(Y_i) < y\right] \quad \text{for} \quad (x,y) \in [0,1]^2,$$

and, accordingly, the *pseudo* tail copula empirical process

$$\begin{aligned}\mathbb{W}_n(x,y) &= \sqrt{n\alpha}\left(\frac{1}{\alpha}\check{C}_n(\alpha x, \alpha y) - \frac{1}{\alpha}C(\alpha x, \alpha y)\right)\\ &=: \sqrt{n\alpha}\left(\check{R}_n(x,y) - R_n(x,y)\right) \quad \text{for} \quad (x,y) \in (0,\infty]^2 \setminus \{\infty,\infty\}.\end{aligned}$$

Below '$\xrightarrow{w}$' denotes weak convergence, $D(I)$ denotes the Skorohod space defined on domain I. Recall that '$\wedge$' denotes the minimum operator and '$\xrightarrow{\mathbb{P}}$' denotes convergence in probability.

Lemma 4.5.1. *Suppose Assumption 4.2.F hold and introduce a weighting function* $q_\eta(t) := t^\eta(1-t)^\eta$, $t > 0$, *with* $\eta \in [0, 1/2)$. *For any fixed* $\alpha \in (0,1)$, *as* $n \to \infty$, *in* $D([0,1/\alpha]^2)$

$$\left\{ \frac{\mathbb{W}_n(x,y)}{q_\eta(\alpha x \wedge \alpha y)},\ x, y \in [0, 1/\alpha] \right\} \xrightarrow{w} \left\{ \frac{1}{\sqrt{\alpha}} \frac{B_C(\alpha x, \alpha y)}{q_\eta(\alpha x \wedge \alpha y)},\ x, y \in [0, 1/\alpha] \right\},$$

where we shall read $\frac{\mathbb{W}_n(x,y)}{q_\eta(\alpha x \wedge \alpha y)} = 0$ *and* $\frac{B_C(\alpha x, \alpha y)}{q_\eta(\alpha x \wedge \alpha y)} = 0$ *if* $x = 0$ *or* $y = 0$ *or* $x = y = 1/\alpha$.

Proof. See Proposition G.1 in Genest and Segers (2009). □

Lemma 4.5.2. *Let* α *be an intermediate sequence such that* $\alpha = \alpha_n \to 0$ *and* $n\alpha_n \to \infty$ *and suppose condition* (4.2.1) *holds. For any* $\eta \in [0, 1/2)$ *and* T *positive, as* $n \to \infty$,

$$\left(\frac{\mathbb{W}_n(x,y)}{(x \wedge y)^\eta}, (x,y) \in (0,T]^2, \frac{\mathbb{W}_n(x,\infty)}{x^\eta}, x \in (0,T], \frac{\mathbb{W}_n(\infty,y)}{y^\eta}, y \in (0,T] \right)$$
$$\xrightarrow{w} \left(\frac{W_R(x,y)}{(x \wedge y)^\eta}, (x,y) \in (0,T]^2, \frac{W_R(x,\infty)}{x^\eta}, x \in (0,T], \frac{W_R(\infty,y)}{y^\eta}, y \in (0,T] \right).$$

in $D((0,T]^2) \times D((0,T]) \times D((0,T])$.

Proof. For convenient presentation, all the limit processes below are defined on the same probability space, via the Skorohod construction. From Lemma 1 in Cai et al. (2015) we know that

$$\left(\frac{\mathbb{W}_n(x,y)}{x^\eta}, (x,y) \in (0,T]^2, \frac{\mathbb{W}_n(x,\infty)}{x^\eta}, x \in (0,T], \frac{\mathbb{W}_n(\infty,y)}{y^\eta}, y \in (0,T] \right)$$
$$\xrightarrow{a.s.} \left(\frac{W_R(x,y)}{x^\eta}, (x,y) \in (0,T]^2, \frac{W_R(x,\infty)}{x^\eta}, x \in (0,T], \frac{W_R(\infty,y)}{y^\eta}, y \in (0,T] \right).$$

Similarly as that in the first coordinate above, we can also show that

$$\left(\frac{\mathbb{W}_n(x,y)}{y^\eta}, (x,y) \in (0,T]^2 \right) \xrightarrow{a.s.} \left(\frac{W_R(x,y)}{y^\eta}, (x,y) \in (0,T]^2 \right).$$

Hence,

$$\sup_{x,y \in (0,T]} \frac{|\mathbb{W}_n(x,y) - W_R(x,y)|}{(x \wedge y)^\eta}$$

$$
\begin{aligned}
&= \max\left\{ \sup_{x,y\in(0,T]} \frac{|\mathbb{W}_n(x,y) - W_R(x,y)|}{x^\eta}, \sup_{x,y\in(0,T]} \frac{|\mathbb{W}_n(x,y) - W_R(x,y)|}{y^\eta} \right\} \\
&\xrightarrow{a.s.} \max\{0,0\} = 0
\end{aligned}
$$

The claim then follows. □

Lemma 4.5.3. *Under the conditions of Theorem 4.2.1 or Theorem 4.2.2, as $n \to \infty$,*

$$
\begin{aligned}
\frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} &= o_{\mathbb{P}}((n\alpha h^2)^{-1/2}), \\
\frac{Y_{n-[n\alpha(1-h))]:n} - Y_{n-[n\alpha(1+h)]:n}}{Q_2(1-\alpha)} &= o_{\mathbb{P}}((n\alpha h^2)^{-1/2}).
\end{aligned}
$$

Proof. We only prove the first statement since the proof of the second one is completely analogous. For any fixed $\alpha \in (0,1)$, by the classical theory of quantile process (c.f., e.g., Example V.12 in Pollard (1984)) and Assumption (4.2.F.a), we have that, for large n,

$$
\begin{aligned}
&X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n} \\
=&Q_1(1-\alpha(1-h)) - Q_1(1-\alpha(1+h)) + O_{\mathbb{P}}(n^{-1/2}) \\
=&O(h) + O_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}((n\alpha h^2)^{-1/2}).
\end{aligned}
$$

When $\alpha = \alpha_n$ is an intermediate sequence, a tail analogue of the above statement can be derived by using, e.g., Theorem 2.4.8 in de Haan and Ferreira (2006) for univariate regularly varying distributions in such a way that

$$
\begin{aligned}
&\frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} \\
=&(1-h)^{-1/\gamma_1} - (1+h)^{-1/\gamma_1} + o_{\mathbb{P}}((n\alpha)^{-1/2}) + o(A(1/\alpha)) \\
=&O(h) + o_{\mathbb{P}}((n\alpha)^{-1/2}) + o(A(1/\alpha)) = o_{\mathbb{P}}((n\alpha h^2)^{-1/2})
\end{aligned}
$$

as $n \to \infty$. □

Lemma 4.5.4. *Under the conditions of Theorem 4.2.1 or Theorem 4.2.2, as $n \to \infty$,*

$$\left(\frac{\widehat{C}(\alpha,\alpha) - \widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)}, \frac{\widehat{ES}_\alpha(X) - \widetilde{ES}_\alpha(X)}{ES_\alpha(X)}, \frac{\widehat{ES}_\alpha(Y) - \widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} \right) = o_{\mathbb{P}}((n\alpha)^{-\frac{1}{2}}). \tag{4.5.1}$$

Proof. The convergence in the first coordinate for fixed α is already noticed in Fermanian et al. (2004), with Lemma 4.5.1 and applications of the delta method; see also, e.g., Lemma 1 in Peng et al. (2012) for details. The proofs for intermediate α is very much the same by using Lemma 4.5.2 instead and we refer to Lemma 1 in Peng and Qi (2010) for more details.

In the following we only prove the convergence in the second coordinate since the proof for the third coordinate is completely analogous. Noting $0 < Q_1(1-\alpha) \le ES_\alpha(X)$, it suffices to show

$$\sqrt{n\alpha} \left(\frac{\widehat{ES}_\alpha(X) - \widetilde{ES}_\alpha(X)}{Q_1(1-\alpha)} \right) \xrightarrow{\mathbb{P}} 0. \tag{4.5.2}$$

Write

$$\begin{aligned}
&\frac{\widehat{ES}_\alpha(X) - \widetilde{ES}_\alpha(X)}{Q_1(1-\alpha)} \\
=&\frac{1}{n\alpha} \sum_{i=1}^{n} \frac{X_i - X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \left(K\left(\frac{1 - \bar{F}_{n1}(X_i)/\alpha}{h} \right) - \mathbb{1}[\bar{F}_{n1}(X_i) < \alpha] \right) \\
&+ \frac{X_{n-\lceil n\alpha \rceil:n}}{Q_1(1-\alpha)} \left(1 - \frac{1}{n\alpha} \sum_{i=1}^{n} \mathbb{1}[\bar{F}_{n1}(X_i)) < \alpha] \right) =: J_1 + J_2.
\end{aligned}$$

Applying Lemma 4.5.3 yields

$$\begin{aligned}
|J_1| \le& \frac{X_{n-[n\alpha(1-h)]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} \cdot \\
&\frac{1}{n\alpha} \sum_{i=1}^{n} \left| K\left(\frac{1 - \bar{F}_{n1}(X_i)/\alpha}{h} \right) - \mathbb{1}[\bar{F}_{n1}(X_i) < \alpha] \right| \\
\le& \frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)} \cdot \frac{1}{n\alpha} \cdot O_{\mathbb{P}}(n\alpha h) = o_{\mathbb{P}}((n\alpha)^{-1/2}).
\end{aligned}$$

The rest is straightforward since

$$|J_2| = \frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\left(1-\frac{\lceil n\alpha\rceil - 1}{n\alpha}\right) = O_{\mathbb{P}}((n\alpha)^{-1}) = o_{\mathbb{P}}((n\alpha)^{-1/2}).$$

□

Proof of Theorem 4.2.1. Write

$$\begin{aligned}\frac{\widetilde{\rho}_\alpha}{\rho_\alpha} - 1 = &\left(\frac{\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1\right)\frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)}\frac{ES_\alpha(Y)}{\widetilde{ES}_\alpha(Y)} + \left(\frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} - 1\right)\frac{ES_\alpha(Y)}{\widetilde{ES}_\alpha(Y)}\\ &- \frac{ES_\alpha(Y)}{\widetilde{ES}_\alpha(Y)}\left(\frac{\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right),\end{aligned} \tag{4.5.3}$$

$$\begin{aligned}\frac{\widehat{\rho}_\alpha}{\rho_\alpha} - 1 = &\left(\frac{\widehat{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1\right)\frac{\widehat{ES}_\alpha(X)}{ES_\alpha(X)}\frac{ES_\alpha(Y)}{\widehat{ES}_\alpha(Y)} + \left(\frac{\widehat{ES}_\alpha(X)}{ES_\alpha(X)} - 1\right)\frac{ES_\alpha(Y)}{\widehat{ES}_\alpha(Y)}\\ &- \frac{ES_\alpha(Y)}{\widehat{ES}_\alpha(Y)}\left(\frac{\widehat{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right).\end{aligned} \tag{4.5.4}$$

Combining these with Lemma 4.5.4 yields that

$$\sqrt{n\alpha}\left(\frac{\widehat{\rho}_\alpha}{\rho_\alpha} - \frac{\widetilde{\rho}_\alpha}{\rho_\alpha}\right) \xrightarrow{\mathbb{P}} 0.$$

For the rest it remains to show that $\sqrt{n\alpha}(\widetilde{\rho}_\alpha/\rho_\alpha - 1) \xrightarrow{d} N(0,\sigma_\alpha^2)$, or, to show that

$$\sqrt{n\alpha}\left(\frac{\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1, \frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} - 1, \frac{\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right) \xrightarrow{d} (\Lambda_\alpha, \Theta_{\alpha,1}, \Theta_{\alpha,2}). \tag{4.5.5}$$

For convenient presentation, all the limit processes below are defined on the same probability space, via the Skorohod construction. However, they are only equal in distribution to the original processes. Using Lemma 4.5.1 with some $\eta \in \left(\frac{1}{2+\delta}, \frac{1}{2}\right)$ we have

$$\left\{\frac{\mathbb{W}_n(x,y)}{q_\eta(\alpha x\wedge\alpha y)},\ x,y\in[0,1/\alpha]\right\} \xrightarrow{a.s.} \left\{\frac{1}{\sqrt{\alpha}}\frac{B_C(\alpha x,\alpha y)}{q_\eta(\alpha x\wedge\alpha y)},\ x,y\in[0,1/\alpha]\right\}. \tag{4.5.6}$$

Applying Vervaat (1972) inverse lemma from (or see Lemma A.0.2 in de Haan and Ferreira (2006)) on the marginal processes $\mathbb{W}_n(\cdot, 1/\alpha)$ and $\mathbb{W}_n(1/\alpha, \cdot)$ around a neighborhood of 1 yields that

$$\begin{aligned}\sqrt{n\alpha}(e_n - 1, e'_n - 1) :=& \sqrt{n\alpha}\left(\check{R}^{\leftarrow}_{n1}(1) - 1, \check{R}^{\leftarrow}_{n2}(1) - 1\right) \\ &\xrightarrow{a.s.} \left(-\frac{1}{\sqrt{\alpha}}B_C(\alpha, 1), -\frac{1}{\sqrt{\alpha}}B_C(1, \alpha)\right), \qquad (4.5.7)\end{aligned}$$

where $\check{R}_{n1}(\cdot) = R_n(\cdot, 1/\alpha) = \frac{1}{\alpha}\check{C}(\alpha\cdot, 1)$ and $\check{R}_{n2}(\cdot) = R_n(1/\alpha, \cdot) = \frac{1}{\alpha}\check{C}(1, \alpha\cdot)$.

Using (4.5.6) and (4.5.7) we then have, cf. pages 52-53 in Gaenssler and Stute (1987),

$$\begin{aligned}&\sqrt{n\alpha}\left(\frac{\check{C}_n(\alpha, \alpha)}{C(\alpha, \alpha)} - 1\right) \\ &\overset{a.s.}{=} \frac{\alpha}{C(\alpha, \alpha)}\left\{\mathbb{W}_n(1, 1) + \dot{C}_1(\alpha, \alpha)\sqrt{n\alpha}\,(e_n - 1) + \dot{C}_2(\alpha, \alpha)\sqrt{n\alpha}\,(e'_n - 1)\right\} + o(1) \\ &\xrightarrow{a.s.} \Lambda_\alpha. \qquad (4.5.8)\end{aligned}$$

Moreover, since we can write

$$\begin{cases} \widetilde{ES}_\alpha(X) \overset{a.s.}{=} -\int_0^{e_n} \check{R}_n(u, \infty) dQ_1(1 - \alpha u) + Q_1(1 - e_n\alpha), \\ ES_\alpha(X) = -\int_0^1 u dQ_1(1 - \alpha u) + Q_1(1 - \alpha), \end{cases}$$

and, therefore,

$$\begin{aligned}&\sqrt{n\alpha}\left(\widetilde{ES}_\alpha(X) - ES_\alpha(X)\right) - ES_\alpha(X)\Theta_{\alpha,1} \\ &\overset{a.s.}{=} -\int_0^1 \left(\mathbb{W}_n(x, 1/\alpha) - \frac{1}{\sqrt{\alpha}}B_C(\alpha x, 1)\right) dQ_1(1 - \alpha x) \\ &\quad + \int_{e_n}^1 \left(\sqrt{n\alpha}(1 - x) + \mathbb{W}_n(x, \infty)\right) dQ_1(1 - \alpha x) \\ &\le \sup_{x\in(0,1)} \frac{\left|\mathbb{W}_n(x, 1/\alpha) - \frac{1}{\sqrt{\alpha}}B_C(\alpha x, 1)\right|}{q_\eta(\alpha x)} \int_0^1 (\alpha x)^\eta dQ_1(1 - \alpha x) \\ &\quad + \sup_{|x-1|\le|e_n-1|} \left|\sqrt{n\alpha}(1 - x) + \mathbb{W}_n(x, \infty)\right| \cdot |Q_1(1 - \alpha e_n) - Q_1(1 - \alpha)| \\ &\xrightarrow{a.s.} 0 + 0 = 0.\end{aligned}$$

Similarly, we also have

$$\sqrt{n\alpha}\left(\widetilde{ES}_\alpha(Y) - ES_\alpha(Y)\right) - ES_\alpha(Y)\Theta_{\alpha,2} \xrightarrow{a.s.} 0.$$

Statement (4.5.5) then follows. □

Proof of Theorem 4.2.2. The proof is analogous to that of Theorem 4.2.1, by replacing $\mathbb{W}_n(x, 1/\alpha)$ by $\mathbb{W}_n(x, \infty)$, $\mathbb{W}_n(1/\alpha, y)$ by $\mathbb{W}_n(\infty, y)$ (since $\bar{F}_1(X_i) < 1$ and $\bar{F}_2(Y_i) < 1$ for all $i = 1, \ldots, n$), and the processes $\frac{1}{\sqrt{\alpha}} B_C(\alpha x, \alpha y)$ by $W_R(x, y)$, $\frac{1}{\sqrt{\alpha}} B_C(\alpha x, 1)$ by $W_R(x, \infty)$, and $\frac{1}{\sqrt{\alpha}} B_C(1, \alpha y)$ by $W_R(\infty, y)$. Particularly, with Lemma 4.5.2 we can show that

$$\sqrt{n\alpha}\left(\frac{\widetilde{C}(\alpha,\alpha)}{C(\alpha,\alpha)} - 1, \frac{\widetilde{ES}_\alpha(X)}{ES_\alpha(X)} - 1, \frac{\widetilde{ES}_\alpha(Y)}{ES_\alpha(Y)} - 1\right) \xrightarrow{d} (\Lambda_0, \Theta_{0,1}, \Theta_{0,2}). \quad (4.5.9)$$

While the proof of the first coordinate-wise convergence is straightforward by recalling the equality in (4.5.8), that of the second and third ones are provided in details in the proof of Proposition 3 in Cai et al. (2015)(as a special case by taking $X = Y$ therein) and thus omitted. The rest follows from (4.5.3), (4.5.4) and Lemma 4.5.4, as in the proof of Theorem 4.2.1. □

To show Theorem 4.2.3, we need some intermediate lemmas for the component-wise jackknife pseudo samples as, for $i = 1, \ldots, n$,

$$\begin{cases} \widehat{V}_{C,i} = n\widehat{C}(\alpha,\alpha) - (n-1)\widehat{C}_i(\alpha,\alpha), \\ \widehat{V}_{X,i} = n\widehat{ES}_\alpha(X) - (n-1)\widehat{ES}_{\alpha,i}(X), \\ \widehat{V}_{Y,i} = n\widehat{ES}_\alpha(Y) - (n-1)\widehat{ES}_{\alpha,i}(Y). \end{cases}$$

Specifically, we shall first develop the joint asymptotics of the jackknife means and jackknife (co)variance based on these component-wise pseudo samples. The (marginal) results below for $\widehat{V}_{C,1}, \ldots, \widehat{V}_{C,n}$ are taken mostly from Peng and Qi (2010) for intermediate α and Peng et al. (2012) for fixed α.

Lemma 4.5.5. *As $n \to \infty$, under the conditions of Theorem 4.2.1 or Theorem 4.2.2,*

$$\sqrt{n\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}-1,\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-1,\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-1\right)\xrightarrow{\mathbb{P}}(0,0,0). \tag{4.5.10}$$

Proof. The convergence in the first coordinate is a direct consequence of Lemma 2 in Peng et al. (2012) and Lemma 2 in Peng and Qi (2010) under the conditions of, respectively, Theorem 4.2.1 and Theorem 4.2.2 by noting that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}-1=&\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\alpha}\widehat{V}_{C,i}-\frac{1}{\alpha}\widehat{C}(\alpha,\alpha)\right\}\left(\frac{C(\alpha,\alpha)}{\alpha}\right)^{-1}\frac{C(\alpha,\alpha)}{\widehat{C}(\alpha,\alpha)}\\=&o_{\mathbb{P}}((n\alpha)^{-1/2}).\end{aligned}$$

In the following we shall only prove the convergence in the second coordinate since the proof for the third coordinate is completely analogous. A crucial step is to observe that

$$\bar{F}_{n1,i}(x)-\bar{F}_{n1}(x)=\frac{1}{n-1}\left(\bar{F}_{n1}(x)-\mathbb{1}[x<X_i]\right)\quad\text{for}\quad x\in\mathbb{R}, \tag{4.5.11}$$

which implies that

$$\sum_{i=1}^{n}\left(\bar{F}_{n1,i}(x)-\bar{F}_{n1}(x)\right)=0\quad\text{for}\quad x\in\mathbb{R}, \tag{4.5.12}$$

$$\max_{1\le i\le n}\sup_{x\in\mathbb{R}}\left|\bar{F}_{n1,i}(x)-\bar{F}_{n1}(x)\right|\le n^{-1}. \tag{4.5.13}$$

Now write

$$\begin{aligned}&\widehat{V}_{X,i}\\&=n\widehat{ES}_\alpha(X)-(n-1)\widehat{ES}_{\alpha,i}(X)\\&=\frac{1}{\alpha}\sum_{j=1}^{n}\left(X_j-X_{n-\lceil n\alpha\rceil:n}\right)\left(K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)-K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\right)\\&+\left\{\frac{1}{\alpha}\left(X_i-X_{n-\lceil n\alpha\rceil:n}\right)K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right)+X_{n-\lceil n\alpha\rceil:n}\right\}\end{aligned}$$

$$=: \widehat{V}_{X,i,1} + \widehat{V}_{X,i,2}. \tag{4.5.14}$$

Applying the mean value theorem yields that, for each pair (i,j), there exists $\varepsilon_{i,j}$ between $\bar{F}_{n1}(X_j)$ and $\bar{F}_{n1,i}(X_j)$ such that

$$\begin{aligned}
&K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\\
=&k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\\
&+\frac{1}{2}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\right)^2.
\end{aligned} \tag{4.5.15}$$

Note that (4.5.13) implies that

$$\sup_{1\le i\le n} \mathbb{1}\left(\left|\frac{1-\varepsilon_{i,j}/\alpha}{h}\right| \le 1\right) \le \mathbb{1}\left(1-h-\frac{1}{n\alpha} \le \frac{\bar{F}_{n1}(X_j)}{\alpha} \le 1+h+\frac{1}{n\alpha}\right). \tag{4.5.16}$$

It follows from (4.5.11) and (4.5.12) that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i,1}\\
=&\frac{1}{n\alpha}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\left(\sum_{i=1}^{n}\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\right)\\
&+\frac{1}{2n\alpha}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}\right)^2\\
=&\frac{1}{2n\alpha}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1}(X_j)-\mathbb{1}[X_j<X_i]}{(n-1)\alpha h}\right)^2\\
=&\frac{1}{2n(n-1)^2\alpha h^2}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1}(X_j)}{\alpha}\right)^2\\
&-\frac{1}{n(n-1)^2\alpha h^2}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\left(\frac{\bar{F}_{n1}(X_j)\mathbb{1}[X_j<X_i]}{\alpha^2}\right)\\
&+\frac{1}{2n(n-1)^2\alpha^2h^2}\sum_{j=1}^{n}\left(X_j - X_{n-\lceil n\alpha\rceil:n}\right)\sum_{i=1}^{n}k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\frac{\mathbb{1}[X_j<X_i]}{\alpha}\\
=:&J_1 - J_2 + J_3.
\end{aligned}$$

We start with the most difficult term J_3. We have that, for some $M > 0$

$$
\begin{aligned}
&\left|\frac{J_3}{Q_1(1-\alpha)}\right| \\
\leq & \frac{M}{n^2\alpha^2h^2}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \mathbb{1}\left(1-h-\frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1+h+\frac{1}{n\alpha}\right) \\
&\cdot \frac{1}{n\alpha}\sum_{i=1}^{n}\mathbb{1}[X_j < X_i] \\
\leq & \frac{M}{n^2\alpha^2h^2}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \\
&\cdot \mathbb{1}\left(1-h-\frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1+h+\frac{1}{n\alpha}\right)\left(\frac{\bar{F}_{n1}(X_j)}{\alpha}\right) \\
\leq & \frac{M}{n^2\alpha^2h^2}\cdot(2n\alpha h+3)\frac{X_{n+1-[n\alpha(1+h)]:n} - X_{n-1-[n\alpha(1-h)]:n}}{Q_1(1-\alpha)}\left(1+h+\frac{1}{n\alpha}\right) \\
= & O((n\alpha h)^{-1})\cdot o_{\mathbb{P}}((n\alpha h^2)^{-1/2}) = o_{\mathbb{P}}((n\alpha)^{-1/2}).
\end{aligned}
$$

Similarly we can show that

$$\left|\frac{J_1}{Q_1(1-\alpha)}\right| = o_{\mathbb{P}}((n\alpha)^{-1/2}), \text{ and } \left|\frac{J_2}{Q_1(1-\alpha)}\right| = o_{\mathbb{P}}((n\alpha)^{-1/2}).$$

Hence,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}\right| \leq \left|\frac{J_1}{Q_1(1-\alpha)}\right| + \left|\frac{J_2}{Q_1(1-\alpha)}\right| + \left|\frac{J_3}{Q_1(1-\alpha)}\right| = o_{\mathbb{P}}((n\alpha)^{-1/2}). \tag{4.5.17}$$

Now, using the mean value theorem (again), we know there exists an $\widetilde{\varepsilon}_{i,j}$ between $\bar{F}_{n1,i}(X_j)$ and $\bar{F}_{n1}(X_j)$ such that

$$
\begin{aligned}
&K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) \\
=&k\left(\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right)\frac{\bar{F}_{n1}(X_j)-\bar{F}_{n1,i}(X_j)}{\alpha h}
\end{aligned} \tag{4.5.18}
$$

and, similar as (4.5.16),

$$\sup_{1\leq i\leq n}\mathbb{1}\left(\left|\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right| \leq 1\right) \leq \mathbb{1}\left(1-h-\frac{1}{n\alpha} \leq \frac{\bar{F}_{n1}(X_j)}{\alpha} \leq 1+h+\frac{1}{n\alpha}\right). \tag{4.5.19}$$

It follows that, for some $M > 0$,

$$
\begin{aligned}
&\frac{1}{Q_1(1-\alpha)}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i,2}-\widehat{ES}_\alpha(X)\right)\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\left(K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right)-K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)\right)\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}k\left(\frac{1-\widetilde{\varepsilon}_{i,i}/\alpha}{h}\right)\frac{\bar{F}_{n1,i}(X_i)-\bar{F}_{n1}(X_i)}{\alpha h}\\
=&\frac{1}{n(n-1)\alpha h}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}k\left(\frac{1-\widetilde{\varepsilon}_{i,i}/\alpha}{h}\right)\frac{\bar{F}_{n1}(X_i)}{\alpha}\\
\leq&\frac{M}{n^2\alpha h}\sum_{i=1}^{n}\frac{\left|X_i-X_{n-\lceil n\alpha\rceil:n}\right|}{Q_1(1-\alpha)}\mathbb{1}\left(1-h-\frac{1}{n\alpha}\leq\frac{\bar{F}_{n1}(X_i)}{\alpha}\leq 1+h+\frac{1}{n\alpha}\right)\\
&\cdot\frac{\bar{F}_{n1}(X_i)}{\alpha}\\
\leq&\frac{M}{n^2\alpha h}(2n\alpha h+3)\frac{X_{n+1-[n\alpha(1+h)]:n}-X_{n-1-[n\alpha(1-h)]:n}}{Q_1(1-\alpha)}\left(1+h+\frac{1}{n\alpha}\right)\\
=&o_{\mathbb{P}}(n^{-1})\cdot o_{\mathbb{P}}((n\alpha h^2)^{-1/2})\cdot O(1)=o_{\mathbb{P}}((n\alpha)^{-1/2}).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\frac{\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i}-\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}=&\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}+\frac{\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{X,i,2}-\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}\\
=&o_{\mathbb{P}}((n\alpha)^{-1/2}).
\end{aligned}
$$

The rest follows from the simple fact that $\frac{Q_1(1-\alpha)}{\widehat{ES}_\alpha(X)}=\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\cdot\frac{ES_\alpha(X)}{\widehat{ES}_\alpha(X)}=O_{\mathbb{P}}(1)$. □

Lemma 4.5.6. *As $n\to\infty$, under the conditions of Theorem 4.2.3*

$$
\max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right|=O_{\mathbb{P}}(1),\quad \max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\right|=o_{\mathbb{P}}((n\alpha)^{1/2}),
$$
$$
\max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|=o_{\mathbb{P}}((n\alpha)^{1/2}).
$$

Proof. From the proofs of Theorem 2 in Peng and Qi (2010) and Theorem 2 in Peng et al. (2012), we have $\max_{1\leq i\leq n}|\widehat{V}_{C,i}|=O_p(1)$ for both intermediate and fixed α. The first claim then follows from the consistency of $\widehat{C}(\alpha,\alpha)/\alpha$ (implied by Theorems 4.2.1 and 4.2.2 above).

Next we shall show the second claim. Recall from (4.5.14) that we can write $\widehat{V}_{X,i} = \widehat{V}_{X,i,1} + \widehat{V}_{X,i,2}$. With (4.5.13) and (4.5.16), we have that, using the Taylor expansion (4.5.15), for some large $M > 0$

$$
\begin{aligned}
&\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}\right| \\
\le&\frac{1}{\alpha h}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\max_{1\le i\le n}\left|\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)\right| \\
&+\frac{1}{2\alpha^2h^2}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right|\max_{1\le i\le n}\left|k'\left(\frac{1-\varepsilon_{i,j}/\alpha}{h}\right)\right| \\
&\cdot\max_{1\le i\le n}\left(\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)\right)^2 \\
\le&\frac{M}{n\alpha h}\sum_{j=1}^{n}\frac{\left|X_j - X_{n-\lceil n\alpha\rceil:n}\right|}{Q_1(1-\alpha)}\mathbb{1}\left(1-h-\frac{1}{n}\le\frac{\bar{F}_{n1}(X_j)}{\alpha}\le 1+h+\frac{1}{n}\right) \\
&+\frac{M}{2n^2\alpha^2h^2}\sum_{j=1}^{n}\frac{\left|X_j - X_{n-\lceil n\alpha\rceil:n}\right|}{Q_1(1-\alpha)}\mathbb{1}\left(1-h-\frac{1}{n}\le\frac{\bar{F}_{n1}(X_j)}{\alpha}\le 1+h+\frac{1}{n}\right) \\
\le&\left\{\frac{M}{n\alpha h}+\frac{M}{2n^2\alpha^2h^2}\right\}\cdot(2n\alpha h+3)\frac{X_{n-[n\alpha(1+h)]:n}-X_{n-[n\alpha(1-h)]:n}}{Q_1(1-\alpha)} \\
=&O(1)\cdot O_{\mathbb{P}}((n\alpha h^2)^{-1/2}) = o_{\mathbb{P}}(1),
\end{aligned}
$$

where for the last line we apply Lemma 4.5.3. It remains to verify that

$$
\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\right| = o_{\mathbb{P}}((n\alpha)^{1/2}). \tag{4.5.20}
$$

Recall from Lemma 11.2 in Owen (2001) that

$$
\max_{1\le i\le n}|X_i| = o_P(n^{1/2}).
$$

When $\alpha\in(0,1)$ is fixed, $\widehat{ES}_\alpha(X)\xrightarrow{\mathbb{P}} ES_\alpha(X) > 0$ and Lemma 11.2 in Owen (2001) yields that

$$
\max_{1\le i\le n}\left|\alpha\widehat{V}_{X,i,2}\right| \le \max_{1\le i\le n}|X_i| + \left|X_{n-\lceil n\alpha\rceil:n}\right|(1+\alpha)
$$

$$=o_{\mathbb{P}}(n^{1/2})+O_{\mathbb{P}}(1)=o_{\mathbb{P}}((n\alpha)^{1/2}).$$

When α is an intermediate sequence, similarly, we have

$$\begin{aligned}\max_{1\leq i\leq n}\left|\alpha\widehat{V}_{X,i,2}\right| \leq &\max_{1\leq i\leq n}|X_i|\mathbb{1}[X_i>X_{n-\lceil n\alpha(1+h)\rceil:n}]+\left|X_{n-\lceil n\alpha\rceil:n}\right|(1+\alpha)\\ \leq &\max\{|X_{n:n}|,\left|X_{n-\lceil n\alpha(1+h)\rceil:n}\right|\}+\left|X_{n-\lceil n\alpha\rceil:n}\right|(1+\alpha)\end{aligned}$$

A fundamental result in extreme value theory (see, e.g., Section 1.1 in de Haan and Ferreira (2006)) tells that

$$X_{n:n}=O_p(Q_1(1-1/n)).$$

Therefore, in conjunction with the regular variation of Q_1 implied by Assumption (4.2.I.a),

$$\begin{aligned}\max_{1\leq i\leq n}\left|\frac{\alpha\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\right| \leq &\frac{\max\{|X_{n:n}|,\left|X_{n-\lceil n\alpha(1+h)\rceil:n}\right|\}}{Q_1(1-\alpha)}+(\alpha+1)\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\\ =&O_{\mathbb{P}}\left(\frac{Q_1(1-1/n)}{Q_1(1-\alpha)}\right)+O_{\mathbb{P}}(1)=o_{\mathbb{P}}((n\alpha)^{\gamma_1})=o_{\mathbb{P}}((n\alpha)^{1/2})\end{aligned}$$

since $\gamma_1\in(0,1/2)$. Recalling $\frac{Q_1(1-\alpha)}{\widehat{ES}_\alpha(X)}=\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\cdot\frac{ES_\alpha(X)}{\widehat{ES}_\alpha(X)}=O_{\mathbb{P}}(1)$, the second part of the lemma follows. The proof for the third part is completely analogous and hence omitted. □

Lemma 4.5.7. *Under Assumptions 4.2.I and 4.2.C, as $\alpha\downarrow 0$,*

$$\Sigma^{(\alpha)}:=Cov(\Lambda_\alpha,\Theta_{\alpha,1},\Theta_{\alpha,2})\to Cov\left(\Lambda_0,\Theta_{0,1},\Theta_{0,2}\right).$$

Proof. It is easy to verify that $\mathrm{Var}(\Lambda_\alpha)\to\mathrm{Var}(\Lambda_0)$. Moreover,

$$\begin{aligned}&\mathrm{Cov}(\Theta_{\alpha,1},\Theta_{\alpha,2})\\ =&\left(\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right)^2\left\{\int_0^1\int_0^1\left(\frac{1}{\alpha}C(\alpha x,\alpha y)\right)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_1(1-\alpha y)}{Q_1(1-\alpha)}\right)\right.\\ &\left.-\alpha\int_0^1\int_0^1 xyd\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_1(1-\alpha y)}{Q_1(1-\alpha)}\right)\right\}\end{aligned}$$

Note that $\sup_{0<x,y\le 1}\frac{\left|\alpha^{-1}C(\alpha x,\alpha y)-R(x,y)\right|}{(x\wedge y)^{\tau}}\to 0$ by assumption, and by Potter's inequality in Potter (1942) (or see Proposition B.1.9 in de Haan and Ferreira (2006)) we have that

$$\sup_{\mathbf{x}\in(0,1)} x^{\tau}\left|\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}-x^{-\gamma_1}\right|\to 0,\text{ and }\sup_{\mathbf{y}\in(0,1)} y^{\tau}\left|\frac{Q_1(1-\alpha y)}{Q_1(1-\alpha)}-y^{-\gamma_1}\right|\to 0. \tag{4.5.21}$$

By some calculations we can therefore show that

$$\begin{aligned}\operatorname{Cov}(\Theta_{\alpha,1},\Theta_{\alpha,2}) \to&(1-\gamma_1)^2\left(\int_0^1\int_0^1 R(x,y)dx^{-\gamma_1}dy^{-\gamma_2}-0\right)\\ =&\operatorname{Cov}(\Theta_{0,1},\Theta_{0,2}),\end{aligned}$$

where we also apply the Karamata's theorem (see, e.g., Theorem B.1.5 in de Haan and Ferreira (2006)) which says that

$$\frac{Q_1(1-\alpha)}{ES_\alpha(X)}=\frac{Q_1(1-\alpha)}{\int_0^1 Q_1(1-\alpha u)du}\to-\gamma_1+1 \tag{4.5.22}$$

Similarly, even easier, we can show $\operatorname{Var}(\Theta_{\alpha,j})\to\operatorname{Var}(\Theta_{0,j})$ for $j=1,2$, and

$$\begin{aligned}&\operatorname{Cov}(\Lambda_\alpha,\Theta_{\alpha,1})\\ =&-\frac{\alpha}{C(\alpha,\alpha)}\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\cdot\\ &\left\{\int_0^1\left(\frac{C(\alpha x,\alpha)}{\alpha}-\dot{C}_1(\alpha,\alpha)x-\dot{C}_2(\alpha,\alpha)\frac{C(\alpha x,\alpha)}{\alpha}\right)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)\right.\\ &\left.-\Big(C(\alpha,\alpha)-\alpha\dot{C}_1(\alpha,\alpha)-\alpha\dot{C}_2(\alpha,\alpha)\Big)\int_0^1 xd\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)\right\}\\ \to&\frac{\gamma_1-1}{R(1,1)}\int_0^1\Big(R(x,1)-\dot{R}_1(1,1)x-\dot{R}_2(1,1)R(x,1)\Big)dx^{-\gamma_1}=\operatorname{Cov}(\Lambda_0,\Theta_{0,1}).\end{aligned}$$

The proof of $\operatorname{Cov}(\Lambda_\alpha,\Theta_{\alpha,2})\to\operatorname{Cov}(\Lambda_0,\Theta_{0,2})$ is completely analogous. □

The following lemma establishes the consistency of the jackknife covariance matrix of the relative estimation errors of component-wise nonparametric estimators, where the smoothing technique plays an important role.

Lemma 4.5.8. *Denote* $\widehat{\mathbf{V}}_i = \left(\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}, \frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}, \frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)$, $i = 1, \ldots, n$. *Under the conditions of Theorem 4.2.1 or Theorem 4.2.2, as* $n \to \infty$,

$$\widehat{\Sigma} - \Sigma^{(\alpha)} := \frac{\alpha}{n} \sum_{i=1}^{n} \left(\widehat{\mathbf{V}}_i - \mathbf{1}\right) \left(\widehat{\mathbf{V}}_i - \mathbf{1}\right)' - \Sigma^{(\alpha)} \xrightarrow{\mathbb{P}} 0.$$

Proof. We only prove the convergence of the (co)variance terms $\widehat{\Sigma}_{1,2}$, $\widehat{\Sigma}_{1,3}$, $\widehat{\Sigma}_{2,3}$, $\widehat{\Sigma}_{2,2}$, and $\widehat{\Sigma}_{3,3}$. The convergence of $\widehat{\Sigma}_{2,1}$, $\widehat{\Sigma}_{3,1}$ and $\widehat{\Sigma}_{3,2}$ then follows by the symmetry of $\widehat{\Sigma}$ (and Σ), and the convergence of $\widehat{\Sigma}_{1,1}$ is readily known by Lemma 3 in Peng and Qi (2010) for intermediate α and Lemma 3 in Peng et al. (2012) for fixed α.

Consistency of $\widehat{\Sigma}_{1,2}$ and $\widehat{\Sigma}_{1,3}$. Recall from (4.5.14) that $\widehat{V}_{X,i} = \widehat{V}_{X,i,1} + \widehat{V}_{X,i,2}$. Using the Taylor expansion (4.5.18), with (4.5.13), (4.5.19) and Lemma 4.5.3, we have that for some large $M > 0$

$$\begin{aligned}
&\frac{\alpha}{n} \sum_{i=1}^{n} \frac{\widehat{V}_{X,i,1}^2}{Q_1^2(1-\alpha)} \\
=&\frac{\alpha}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\right. \\
&\left.\cdot \left(K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\right)\right)^2 \\
\le&\alpha \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)^2 \\
&\cdot \left(K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\right)^2 \\
=&\frac{1}{\alpha h^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)^2 k^2\left(\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right)\left(\bar{F}_{n1,i}(X_j) - \bar{F}_{n1}(X_j)\right)^2 \\
\le&\frac{M}{n\alpha h^2} \sum_{j=1}^{n} \left(\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)^2 \mathbb{1}\left(1 - h - \frac{1}{n} \le \frac{\bar{F}_{n1}(X_j)}{\alpha} \le 1 + h + \frac{1}{n}\right) \\
\le&\frac{M}{n\alpha h^2}(2n\alpha h + 3) \cdot \left(\frac{X_{n-[n\alpha(1-h))]:n} - X_{n-[n\alpha(1+h)]:n}}{Q_1(1-\alpha)}\right)^2 = o_{\mathbb{P}}((n\alpha)^{-1}) = o_{\mathbb{P}}(1).
\end{aligned}$$

Write $\widehat{V}_{C,i} = \widehat{V}_{C,i,1} + \widehat{V}_{C,i,2}$, where

$$\begin{aligned}\widehat{V}_{C,i,1} =& \sum_{j=1}^{n} \left\{ K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)\right.\\ &\left. - K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2,i}(Y_j)/\alpha}{h}\right)\right\},\\ \widehat{V}_{C,i,2} =& K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2,i}(Y_i)/\alpha}{h}\right)\end{aligned}$$

We have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} \widehat{V}_{C,i,2} \cdot \frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\\ =&\frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right) K^2\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2,i}(Y_i)/\alpha}{h}\right)\\ &+\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{1-\bar{F}_{n1,i}(X_i)/\alpha}{h}\right) K\left(\frac{1-\bar{F}_{n2,i}(Y_i)/\alpha}{h}\right)\\ =&\frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\mathbb{1}\left(X_i > X_{n-\lceil n\alpha\rceil:n}, Y_i > Y_{n-\lceil n\alpha\rceil:n}\right)\\ &+\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\widehat{C}(\alpha,\alpha) + o_{\mathbb{P}}(1)\\ =&\int_{e_n}^{0}\check{R}_n(u, e_n')d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right) + C(\alpha,\alpha) + o_{\mathbb{P}}(1)\\ =&-\int_0^1 R_n(u,1)d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right) + C(\alpha,\alpha) + o_{\mathbb{P}}(1),\end{aligned}$$

where in the last step we recall from the proofs of Theorems 4.2.1 and 4.2.2 that $e_n := R_{n1}^{\leftarrow}(1) = 1 + o_{\mathbb{P}}(1)$ and $e_n' := R_{n2}^{\leftarrow}(1) = 1 + o_{\mathbb{P}}(1)$.

Moreover, similarly as the proof of (4.5.17), we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,1}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\\ =&\frac{1}{n\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(\frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right) K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right) + \alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right\}\\ &\cdot\frac{\bar{F}_{n2,i}(Y_j) - \bar{F}_{n2}(Y_j)}{\alpha h} K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right) k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(\frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)+\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right\}\\
&\cdot\frac{\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)}{\alpha h}k\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)+o_{\mathbb{P}}(1)\\
=:\ & T_1+T_2+o_{\mathbb{P}}(1).
\end{aligned}
$$

Note that $\bar{F}_{n2,i}(y)-\bar{F}_{n2}(y)=\frac{1}{n-1}\left(\bar{F}_{n2}(y)-\mathbb{1}\left[\bar{F}_{n2}(Y_i)<\bar{F}_{n2}(y)\right]\right)$. Therefore, in conjunction with (4.5.5), (4.5.9) and (4.5.22), we have

$$
\begin{aligned}
&T_1\\
=&\frac{\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}\frac{1}{(n-1)\alpha h}\sum_{j=1}^{n}\bar{F}_{n2}(Y_j)k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\\
-&\frac{1}{(n-1)\alpha h}\sum_{j=1}^{n}\left\{\frac{1}{n\alpha}\sum_{i=1}^{n}\frac{X_i-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)\right.\\
&\left.\mathbb{1}\left[\bar{F}_{n2}(Y_i)<\bar{F}_{n2}(Y_j)\right]\right\}k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\\
-&\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{1}{(n-1)\alpha h}\sum_{j=1}^{n}\left\{\frac{1}{n\alpha}\sum_{i=1}^{n}\mathbb{1}\left[\bar{F}_{n2}(Y_i)<\bar{F}_{n2}(Y_j)\right]\right\}\\
&\cdot k\left(\frac{1-\bar{F}_{n2}(Y_j)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)\\
=&\ \dot{C}_2(\alpha,\alpha)\frac{ES_\alpha(X)}{Q_1(1-\alpha)}\alpha+\dot{C}_2(\alpha,\alpha)\int_0^1 R_n(u,1)d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)\\
&-\alpha\dot{C}_2(\alpha,\alpha)+o_{\mathbb{P}}(1)\\
=&\ \alpha\dot{C}_2(\alpha,\alpha)\left\{\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-1\right\}+\dot{C}_2(\alpha,\alpha)\int_0^1 R_n(u,1)d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)\\
&+o_{\mathbb{P}}(1)
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
T_2&=\dot{C}_1(\alpha,\alpha)\left\{\alpha\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-\frac{ES_\alpha(X)}{Q_1(1-\alpha)}+1-\alpha\right\}+o_{\mathbb{P}}(1)\\
&=-\dot{C}_1(\alpha,\alpha)(1-\alpha)\left\{\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-1\right\}+o_{\mathbb{P}}(1).
\end{aligned}
$$

To conclude we have, again recalling (4.5.5), (4.5.9) and (4.5.22),

$$
\begin{aligned}
&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\\
=&\frac{\alpha}{\widehat{C}(\alpha,\alpha)}\frac{Q_1(1-\alpha)}{\widehat{ES}_\alpha(X)}\left\{\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i}\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}+\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,1}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\right.\\
&\left.+\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{C,i,2}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\right\}\\
=&-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\int_0^1\frac{R_n(u,1)}{R_n(1,1)}d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)\\
&-\dot{C}_1(\alpha,\alpha)\frac{1}{R_n(1,1)}\left\{1-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right\}\\
&+\dot{C}_2(\alpha,\alpha)\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\int_0^1\frac{R_n(u,1)}{R_n(1,1)}d\left(\frac{Q_1(1-\alpha u)}{Q_1(1-\alpha)}\right)\\
&-\alpha\left(1-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right)\left\{1-\frac{\dot{C}_1(\alpha,\alpha)}{R_n(1,1)}-\frac{\dot{C}_2(\alpha,\alpha)}{R_n(1,1)}\right\}+\alpha+o_{\mathbb{P}}(1)\\
=&\Sigma_{1,2}^{(\alpha)}+\alpha+o_{\mathbb{P}}(1).
\end{aligned}
$$

It follows that, in conjunction with Lemma 4.5.5, as $n\to\infty$

$$
\begin{aligned}
\widehat{\Sigma}_{1,2}=&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}+\alpha\\
=&\Sigma_{1,2}^{(\alpha)}+\alpha-\alpha-\alpha+\alpha+o_{\mathbb{P}}(1)=\Sigma_{1,2}^{(\alpha)}+o_{\mathbb{P}}(1).
\end{aligned}
$$

Similarly, we can show that $\widehat{\Sigma}_{1,3}=\Sigma_{1,3}^{(\alpha)}+o_{\mathbb{P}}(1)$.

Consistency of $\widehat{\Sigma}_{2,3}$, $\widehat{\Sigma}_{2,2}$ and $\widehat{\Sigma}_{2,2}$. Next, we write $\widehat{V}_{X,i}=\widehat{V}_{X,i,1}+\widehat{V}_{X,i,2}$ as defined in equation (4.5.14). Analogously we define $\widehat{V}_{Y,i,1}$ and $\widehat{V}_{Y,i,2}$. With (4.5.18) and (4.5.19), there exists some $M>0$ such that for all $i=1,\dots,n$

$$
\begin{aligned}
&\left|\frac{\widehat{V}_{X,i,1}}{Q_1(1-\alpha)}\right|\\
\leq&\sum_{j=1}^{n}\left|\frac{X_j-X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right|\left|K\left(\frac{1-\bar{F}_{n1}(X_j)/\alpha}{h}\right)-K\left(\frac{1-\bar{F}_{n1,i}(X_j)/\alpha}{h}\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
=&\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| k\left(\frac{1-\widetilde{\varepsilon}_{i,j}/\alpha}{h}\right)\frac{\left|\bar{F}_{n1,i}(X_j)-\bar{F}_{n1}(X_j)\right|}{\alpha h}\\
\leq&\frac{M}{n\alpha h}\sum_{j=1}^{n}\left|\frac{X_j - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right| \mathbb{1}\left(1-h-\frac{1}{n\alpha}\leq\frac{\bar{F}_{n1}(X_j)}{\alpha}\leq 1+h+\frac{1}{n\alpha}\right)\\
\leq&\frac{M}{n\alpha h}(2n\alpha h+3)\frac{X_{n-[n\alpha(1-\alpha/h)]:n}-X_{n-[n\alpha(1+\alpha/h)]:n}}{Q_1(1-\alpha)}.
\end{aligned}
$$

It follows that, together with Lemma 4.5.3,

$$
\begin{aligned}
\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{X,i,1}}{ES_\alpha(X)}\right)^2\leq&\frac{\alpha M^2}{n^2\alpha^2h^2}(2n\alpha h+3)^2\left(\frac{X_{n-[n\alpha(1-\alpha/h)]:n}-X_{n-[n\alpha(1+\alpha/h)]:n}}{Q_1(1-\alpha)}\right)^2\\
=&O(\alpha)\cdot o_{\mathbb{P}}((n\alpha h^2)^{-1/2})=o_{\mathbb{P}}(1).
\end{aligned}
$$

Similarly, we also have $\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{Y,i,1}}{ES_\alpha(X)}\right)^2=o_{\mathbb{P}}(1)$. Moreover, we have

$$
\begin{aligned}
&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,2}}{Q_1(1-\alpha)}\frac{\widehat{V}_{Y,i,2}}{Q_2(1-\alpha)}\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\left(\frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\left(\frac{Y_i - Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\right)\\
&\cdot K\left(\frac{1-\bar{F}_{n1}(X_i)/\alpha}{h}\right)K\left(\frac{1-\bar{F}_{n2}(Y_i)/\alpha}{h}\right)+\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{\widehat{ES}_\alpha(Y)}{Q_2(1-\alpha)}\\
&+\alpha\frac{Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\frac{\widehat{ES}_\alpha(X)}{Q_1(1-\alpha)}-\alpha\frac{X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\frac{Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\\
=&\frac{1}{n\alpha}\sum_{i=1}^{n}\left(\frac{X_i - X_{n-\lceil n\alpha\rceil:n}}{Q_1(1-\alpha)}\right)\left(\frac{Y_i - Y_{n-\lceil n\alpha\rceil:n}}{Q_2(1-\alpha)}\right)\\
&\cdot\mathbb{1}\left[X_j>X_{n-\lceil n\alpha\rceil:n},Y_j>Y_{n-\lceil n\alpha\rceil:n}\right]+\frac{\alpha ES_\alpha(Y)}{Q_2(1-\alpha)}+\frac{\alpha ES_\alpha(X)}{Q_1(1-\alpha)}-\alpha+o_{\mathbb{P}}(1)\\
=&\int_0^{e_n}\int_0^{e_n'}\check{R}_n(x,y)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_2(1-\alpha y)}{Q_2(1-\alpha)}\right)+\alpha\frac{ES_\alpha(Y)}{Q_2(1-\alpha)}\\
&+\alpha\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-\alpha+o_{\mathbb{P}}(1)\\
=&\int_0^{1}\int_0^{1}R_n(x,y)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_2(1-\alpha y)}{Q_2(1-\alpha)}\right)+\alpha\frac{ES_\alpha(Y)}{Q_2(1-\alpha)}\\
&+\alpha\frac{ES_\alpha(X)}{Q_1(1-\alpha)}-\alpha+o_{\mathbb{P}}(1),
\end{aligned}
$$

where in the last step we apply Lemma 4.5.1 for fixed α with $\eta \in \left(\frac{1}{2+\delta}, \frac{1}{2}\right)$ and Lemma 4.5.2 for intermediate α with $\eta \in (\gamma_1, 1/2)$.

It follows that, again recalling (4.5.5), (4.5.9) and (4.5.22),

$$\begin{aligned}
&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}\\
=&\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\frac{Q_2(1-\alpha)}{ES_\alpha(Y)}\int_0^1\int_0^1 R_n(x,y)d\left(\frac{Q_1(1-\alpha x)}{Q_1(1-\alpha)}\right)d\left(\frac{Q_2(1-\alpha y)}{Q_2(1-\alpha)}\right)\\
&-\alpha\left(1-\frac{Q_1(1-\alpha)}{ES_\alpha(X)}\right)\left(1-\frac{Q_2(1-\alpha)}{ES_\alpha(Y)}\right)+\alpha+o_{\mathbb{P}}(1)\\
=&\Sigma_{2,3}^{(\alpha)}+\alpha+o_{\mathbb{P}}(1).
\end{aligned}$$

Analogously, we can also show that

$$\begin{aligned}
\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\right)^2 &= \Sigma_{2,2}^{(\alpha)}+\alpha+o_{\mathbb{P}}(1),\\
\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}\right)^2 &= \Sigma_{3,3}^{(\alpha)}+\alpha+o_{\mathbb{P}}(1).
\end{aligned}$$

Hence, in conjunction with Lemma 4.5.5,

$$\begin{aligned}
&\widehat{\Sigma}_{2,3}\\
=&\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{Y,i,2}}{\widehat{ES}_\alpha(Y)}-\frac{\alpha}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{X,i,2}}{\widehat{ES}_\alpha(X)}+\alpha+o_{\mathbb{P}}(1)\\
=&\Sigma_{2,3}^{(\alpha)}+\alpha-\alpha-\alpha+\alpha+o_{\mathbb{P}}(1)=\Sigma_{2,3}^{(\alpha)}+o_{\mathbb{P}}(1).
\end{aligned}$$

Similarly, $\widehat{\Sigma}_{2,2}=\Sigma_{2,2}^{(\alpha)}+o_{\mathbb{P}}(1)$ and $\widehat{\Sigma}_{3,3}=\Sigma_{3,3}^{(\alpha)}+o_{\mathbb{P}}(1)$. □

Finally, we combine the above component-wise results to establish the asymptotics of the jackknife pseudo sample of our relative risk measure, that is, $\widehat{V}_{\rho,1},\ldots,\widehat{V}_{\rho,n}$.

Lemma 4.5.9. *As $n\to\infty$, under the conditions of Theorem 4.2.1 or Theorem 4.2.2*

$$\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}\right| = o_{\mathbb{P}}((n\alpha)^{1/2}),\qquad \sqrt{n\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}-1\right)\xrightarrow{\mathbb{P}}0,$$

$$\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}-1\right)^2-\sigma_\alpha^2 \xrightarrow{\mathbb{P}} 0.$$

Proof. Note that we can write, for $i=1,\dots,n$,

$$\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}=\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}+\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}+\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha},$$

and therefore

$$\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}=\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}+\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}. \tag{4.5.23}$$

Hence,

$$\left(1-\frac{1}{n\alpha}\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|\right)\left|\frac{\alpha\widehat{V}_i}{\widehat{\rho}_\alpha}\right|\le\left|\frac{\alpha\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right|+\left|\frac{\alpha\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\right|+\left|\frac{\alpha\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right|$$
$$+\frac{1}{n\alpha}\left|\frac{\alpha\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right|\left|\frac{\alpha\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\right|.$$

Lemma 4.5.6 implies that

$$\left(1-o_{\mathbb{P}}((n\alpha)^{-1/2})\right)\max_{1\le i\le n}\left|\frac{\alpha\widehat{V}_{\rho,i}}{\widehat{\rho}_\alpha}\right|=o_{\mathbb{P}}((n\alpha)^{1/2}),$$

and then the *first* claim follows.

Note that (4.5.23) also yields that

$$\frac{\widehat{V}_{\rho,i}}{\widehat{\rho}_i}-1=\left(\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}-1\right)+\left(\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}-1\right)-\left(\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-1\right)$$
$$+\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-1}T_i \tag{4.5.24}$$

where

$$T_i=\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}+\frac{1}{n}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}-\frac{1}{n}\left(\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^2$$
$$-\frac{1}{n}\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{X,i}}{\widehat{ES}_\alpha(X)}=:T_{i,1}+T_{i,2}-T_{i,3}-T_{i,4}.$$

With Lemmas 4.5.5 and 4.5.8, applying Cauchy-Schwarz inequality yields that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}|T_{i,1}| =&\frac{1}{n^2}\sum_{i=1}^{n}\left|\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right| \\ \leq&\frac{1}{n\alpha}\sqrt{\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{C,i}}{\widehat{C}(\alpha,\alpha)}\right)^2\cdot\frac{\alpha}{n}\sum_{i=1}^{n}\left(\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^2} = O_{\mathbb{P}}((n\alpha)^{-1}).\end{aligned}$$

Analogously, we can also show that for all $j=2,3,4$, $\frac{1}{n}\sum_{i=1}^{n}|T_{i,j}| = O_{\mathbb{P}}\left((n\alpha)^{-1}\right)$. Therefore

$$\frac{1}{n}\sum_{i=1}^{n}|T_i| \leq \sum_{j=1}^{4}\frac{1}{n}\sum_{i=1}^{n}|T_{i,j}| = O_{\mathbb{P}}((n\alpha)^{-1}).$$

Recalling $\max_{1\leq i\leq n}\left|\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right| = o_{\mathbb{P}}((n\alpha)^{-1/2})$, we have that

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-1}T_i\right| \leq&\frac{1}{n}\sum_{i=1}^{n}|T_i|\max_{1\leq i\leq n}\left|\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-1}\right| \\ =&O_{\mathbb{P}}((n\alpha)^{-1})\cdot O_{\mathbb{P}}(1) = o_{\mathbb{P}}((n\alpha)^{-1/2}),\end{aligned}$$

and then the *second* claim follows by applying Lemma 4.5.5 with (4.5.24).

Similarly, we have

$$\begin{aligned}&\frac{\alpha}{n}\sum_{i=1}^{n}\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-2}T_i^2 \\ \leq&\frac{\alpha}{n}\sum_{i=1}^{n}T_i^2\max_{1\leq i\leq n}\left|\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-2}\right| \\ \leq&n\alpha\left(\frac{1}{n}\sum_{i=1}^{n}|T_i|\right)^2\max_{1\leq i\leq n}\left|\left(1-\frac{1}{n}\frac{\widehat{V}_{Y,i}}{\widehat{ES}_\alpha(Y)}\right)^{-2}\right| \\ =&n\alpha\cdot O_{\mathbb{P}}\left((n\alpha)^{-2}\right)\cdot O_{\mathbb{P}}(1) = O_{\mathbb{P}}\left((n\alpha)^{-1}\right) = o_{\mathbb{P}}(1).\end{aligned}$$

The *third* claim then follows from Lemma 4.5.8, using (4.5.24) again. □

Proof of Theorem 4.2.3. Set $m_i = m_i(\rho_\alpha) = \frac{\alpha\widehat{V}_{\rho,i}}{\rho_\alpha} - \alpha$, $i=1,\ldots,n$, $m_n^* = \max_{1\leq i\leq n}|m_i|$, $\bar{m}_n = n^{-1}\sum_{i=1}^{n}m_i$, $S_n = n^{-1}\sum_{i=1}^{n}m_i^2$. Now Lemma 4.5.7

and Lemma 4.5.9 in conjunction with Theorem 4.2.1 and Theorem 4.2.2 imply that, as $n \to \infty$,

$$m_n^* = o_{\mathbb{P}}\left((n\alpha)^{1/2}\right), \quad \sqrt{\frac{n}{\alpha}} \frac{\bar{m}_n}{\sigma_\alpha} \xrightarrow{d} N(0,1), \quad \text{and} \quad \frac{S_n}{\alpha\sigma_\alpha^2} \xrightarrow{\mathbb{P}} 1. \tag{4.5.25}$$

When taking $\theta = \rho_\alpha$, equation (4.2.2) can be rewritten as

$$p_i = \frac{1}{n} \frac{1}{1 + \widetilde{\lambda} m_i}$$

with $\widetilde{\lambda} = \lambda\rho_\alpha/\alpha$ and equation (4.2.3) can be rewritten as

$$\frac{1}{n} \sum_{i=1}^{n} \frac{m_i}{1 + \widetilde{\lambda} m_i} = 0.$$

Now, following the proof of Theorem 2 in Peng and Qi (2010), statement (4.5.25) implies that $\widetilde{\lambda} = O_{\mathbb{P}}((n\alpha)^{-1/2})$ and, furthermore,

$$\widetilde{\lambda} = S_n^{-1} \bar{m}_n + o_{\mathbb{P}}((n\alpha)^{-1/2}).$$

Hence, by a Taylor expansion and again (4.5.25), we have, as $n \to \infty$,

$$\begin{aligned} -2 \log \widehat{\mathcal{R}}(\rho_\alpha) &= 2 \sum_{i=1}^{n} \widetilde{\lambda} m_i - \sum_{i=1}^{n} \widetilde{\lambda}^2 m_i^2 + o_{\mathbb{P}}(1) \\ &= n S_n^{-1} \bar{m}_n^2 + o_{\mathbb{P}}(1) \\ &\xrightarrow{d} \chi^2(1). \end{aligned}$$

□

Bibliography

Agarwal, V., S. Ruenzi, and F. Weigert (2016). Tail risk in hedge funds: A unique view from portfolio holdings. *Journal of Financial Economics*. Forthcoming.

Alexander, K. S. (1987). Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probability Theory and Related Fields 75*, 379–423.

Arcones, M. A., H. Cui, and Y. Zuo (2006). Empirical depth processes. *TEST 15*, 151–177.

Basel Committee on Banking Supervision (2005). Basel ii: International convergence of capital measurement and capital standards: A revised framework. November.

Baysal, R. E. and J. Staum (2008). Empirical likelihood for value-at-risk and expected shortfall. *Journal of Risk 11*, 3–32.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2006). *Statistics of extremes: theory and applications*. Chichester: Wiley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307 – 327.

Cai, J.-J., J. H. J. Einmahl, and L. de Haan (2011). Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics 39*, 1803–1826.

Cai, J.-J., J. H. J. Einmahl, L. Haan, and C. Zhou (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*, 417–442.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association 91*, 862–872.

Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics 6*, 87–107.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance 1*, 223–236.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: an Introduction.* New York: Springer.

de Haan, L. and S. Resnick (1979). Derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. *Stochastic Processes and their Applications 8*, 349 – 355.

de Haan, L. and S. Resnick (1987). On regular variation of probability densities. *Stochastic Processes and their Applications 25*, 83–93.

Dekkers, A. L. M., J. H. J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics 17*, 1833–1855.

Donoho, D. L. and M. Gasko (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics 20*, 1803–1827.

Drees, H. and X. Huang (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis 64*, 25 – 46.

Dudley, R. M. (2002). *Real analysis and probability.* Cambridge University

Press.

Einmahl, J. H. J. (1997). Poisson and gaussian approximation of weighted local empirical processes. *Stochastic Processes and their Applications 70*, 31 – 58.

Einmahl, J. H. J., L. de Haan, and D. Li (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics 34*, 1987–2014.

Einmahl, J. H. J. and A. Krajina (2016). Empirical likelihood based testing for multivariate regular variation. In preparation.

Einmahl, J. H. J., J. Li, and R. Y. Liu (2015a). Bridging centrality and extremity: Refining empirical data depth using extreme value statistics. *The Annals of Statistics 43*, 2738–2765.

Einmahl, J. H. J., J. Li, and R. Y. Liu (2015b). Bridging centrality and extremity: refining empirical data depth using extreme value statistics. *The Annals of Statistics 43*, 2738–2765.

Fermanian, J.-D., D. Radulovic, and M. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli 10*, 847–860.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics 1*, 255–293.

Gaenssler, P. and W. Stute (1987). *Seminar on empirical processes.* Basel: Springer.

Gänssler, P. (1983). Empirical processes. In S. S. Gupta (Ed.), *Lecture Notes-Monograph Series*, Volume 3, Hayward, CA. Institute of Mathematical Statistics.

Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics 37*, 2990–3022.

Girard, S. and G. Stupfler (2014, September). Asymptotic behaviour of extreme geometric quantiles and their estimation under moment conditions. Working Paper.

Gong, Y., L. Peng, and Y. Qi (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis 101*, 1520 – 1531.

Graham, J. R. and C. R. Harvey (1996). Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. *Journal of Financial Economics 42*, 397 – 421.

Graham, J. R. and C. R. Harvey (1997). Grading the performance of market-timing newsletters. *Financial Analysts Journal 53*, 54–66.

Hallin, M., D. Paindaveine, and M. Šiman (2010). Multivariate quantiles and multiple-output regression quantiles: From l_1 optimization to halfspace depth. *The Annals of Statistics 38*, 635–669.

Hashorva, E. (2006). On the regular variation of elliptical random vectors. *Statistics & Probability Letters 76*, 1427 – 1434.

He, X. and G. Wang (1997). Convergence of depth contours for multivariate datasets. *The Annals of Statistics 25*, 495–504.

He, Y. and J. H. J. Einmahl (2016). Estimation of extreme depth-based quantile regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. DOI: 10.1111/rssb.12163.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*, 1163–1174.

Hoffmann-Jørgensen, J. (1991). *Stochastic Processes on Polish Spaces.* Aarhus Universitet. Matematisk Institut.

Jing, B.-Y., J. Yuan, and W. Zhou (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association 104*, 1224–1232.

Kelly, B. and H. Jiang (2014). Tail risk and asset prices. *Review of Financial Studies 27*, 2841–2871.

Kong, L. and I. Mizera (2012). Quantile tomography: Using quantiles with multivariate data. *Statistica Sinica 22*, 1589–1610.

Liu, R. Y. (1992). Data depth and multivariate rank tests. In Y. Dodge (Ed.), *L1-Statistical Analysis and Related Methods*, Amsterdam, pp. 279–294. North-Holland.

Liu, R. Y., J. M. Parelius, and K. Singh (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by liu and singh). *The Annals of Statistics 27*, 783–858.

Longin, F. and B. Solnik (2001). Extreme correlation of international equity markets. *The Journal of Finance 56*, 649–676.

Luo, X. and P. V. Shevchenko (2010). The t copula with multiple parameters of degrees of freedom: bivariate characteristics and application to risk management. *Quantitative Finance 10*, 1039–1054.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, Volume 2, pp. 49–55.

Massé, J.-C. (2004). Asymptotics for the tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli 10*, 397–419.

McNeil, A. J. and A. D. Smith (2012). Multivariate stress scenarios and solvency. *Insurance: Mathematics and Economics 50*, 299–308.

Modigliani, F. and L. Modigliani (1997). Risk-adjusted performance. *The Journal of Portfolio Management 23*, 45–54.

Mosteller, F. and J. W. Tukey (1977). *Data analysis and regression: a second course in statistics*. Reading, MA: Addison-Wesley.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*, 347–370.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econo-*

metrica 55, 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*, 631–653.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*, 90–120.

Owen, A. B. (2001). *Empirical likelihood*. New York: Chapman and Hall.

Peng, L. and Y. Qi (2010). Smoothed jackknife empirical likelihood method for tail copulas. *TEST 19*, 514–536.

Peng, L., Y. Qi, and I. Van Keilegom (2012). Jackknife empirical likelihood method for copulas. *TEST 21*, 74–92.

Peng, L., Y. Qi, R. Wang, and J. Yang (2012). Jackknife empirical likelihood method for some risk measures and related quantities. *Insurance: Mathematics and Economics 51*, 142 – 150.

Peng, L., X. Wang, and Y. Zheng (2015). Empirical likelihood inference for Haezendonck-Goovaerts risk measure. *European Actuarial Journal 5*, 427–445.

Pollard, D. (1984). *Convergence of stochastic processes*. New York: Springer.

Poon, S.-H., M. Rockinger, and J. Tawn (2004). Extreme value dependence in financial markets: diagnostics, models, and financial implications. *Review of Financial Studies 17*, 581–610.

Potter, H. (1942). The mean values of certain dirichlet series, ii. *Proceedings of the London Mathematical Society 2*, 1–19.

Resnick, S. I. (2007). *Extreme Values, Regular Variation, and Point Processes*. Springer.

Serfling, R. (2002). A depth function and a scale curve based on spatial quantiles. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods (Neuchâtel, 2002)*, pp. 25–38. Basel: Birkäuser.

Serfling, R. (2006). Depth functions in nonparametric multivariate inference. In R. Liu and R. Serfling (Eds.), *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Volume 72, pp. 1–16. Providence, RI: American Mathematical Society.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *Journal of Nonparametric Statistics 22*, 915–936.

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Sibuya, M. (1959). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics 11*, 195–210.

Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics 15*, 1174–1207.

Struyf, A. J. and P. J. Rousseeuw (1999). Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis 69*, 135–153.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974)*, Volume 2, Montreal, pp. 523–531. Canad. Math. Congress.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer.

van Oordt, M. R. C. and C. Zhou (2016). Systematic tail risk. *Journal of Financial and Quantitative Analysis 51*, 685–705.

Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Probability Theory and Related Fields 23*, 245–253.

Wang, X. and L. Peng (2016). Inference for intermediate haezen-donck–goovaerts risk measure. *Insurance: Mathematics and Economics 68*, 231 – 240.

Weissman, I. (1978). Estimation of parameters and larger quantiles based on the k largest observations. *Journal of the American Statistical Association 73*, 812–815.

Yeh, A. B. and K. Singh (1997). Balanced confidence regions based on tukey's depth and the bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59*, 639–652.

Zuo, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics 31*, 1460–1490.

Zuo, Y. and R. Serfling (2000a). General notions of statistical depth function. *The Annals of Statistics 28*, 461–482.

Zuo, Y. and R. Serfling (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *The Annals of Statistics 28*, 483–499.